2016

The Genomics of Speciation within the Globally Distributed Blue-Winged Ducks

Joel Thomas Nelson
Wright State University

THE GENOMICS OF SPECIATION WITHIN THE GLOBALLY DISTRIBUTED BLUE-WINGED DUCKS

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science

By

Joel T. Nelson
B.S., Salisbury University, 2013
B.S., University of Maryland Eastern Shore 2013

2016
Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

May 18, 2016

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER
MY SUPERVISION BY Joel T. Nelson ENTITLED The genomics of speciation within the globally distributed blue-winged ducks.
BE ACCEPTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

_____
Jeffrey L. Peters, Ph.D.
Thesis Director

_____
David Goldstein, Ph.D.
Chair, Department of Biological Sciences

Committee on Final Examination

_____
Jeffrey L. Peters, Ph.D.

_____
Dan E. Krane, Ph.D.

_____
John O. Stireman, Ph.D.

_____
Robert E. W. Fyffe, Ph.D.
Vice President for Research and
Dean of the Graduate School

ABSTRACT

Nelson, Joel T. M.S. Department of Biological Sciences, Wright State University, 2016.
The Genomics of Speciation within the Globally Distributed Blue-Winged Ducks

The ability to disperse over long distances can result in a high propensity for colonizing new geographic regions and lead to lineage diversification through allopatric speciation and divergence-with-gene-flow. Here, I use the combination of DNA sequences and coalescent methods to examine gene flow, and phylogenetic relationships within cosmopolitan blue-winged duck lineage (genus *Anas*). I found two primary sub-lineages, the globally-distributed shoveler group and the New World blue-winged/cinnamon teal group. However, I found evidence of gene flow from the migratory Holarctic northern shoveler (*A. clypeata*) and the more sedentary African Cape shoveler (*A. smithii*) into the Australasian shoveler (*A. rhynchotis*), supporting my hypothesis of intercontinental gene flow. Given the diverse mechanisms of speciation within this complex, the shovelers and blue-winged/cinnamon teals can serve as an effective model system for examining how the genome diverges under different evolutionary processes and how genetic variation is partitioned among highly dispersive taxa.

TABLE OF CONTENTS

PAGE

LIST OF FIGURES

FIGURE PAGE

LIST OF TABLES

TABLE PAGE

ACKNOWLEDGEMENTS

This work was supported by a Start-up Grant and a Research Initiation Grant from the College of Science and Mathematics at Wright State University to J.L. Peters. We thank the University of Washington Burke Museum and the Louisiana State University Museum of Natural Science for tissue loans. Additional sample collection was supported by grants from the Institute of Arctic Biology, Alaska EPSCoR (NSF EPS-0092040 & EPS-0346770) and NSF (DEB-0444748) to K.G. McCracken, by USAID via subcontract to the Wildlife Conservation Society's GAINS programme to G. Cumming, grants from the Delta Waterfowl Foundation and Frank M.

Chapter 1. Divergence and gene flow in the globally distributed blue-winged ducks

INTRODUCTION

Speciation is a dynamic process where interactions among evolutionary forces, such as natural selection, genetic drift and gene flow, result in variable rates of genetic divergence. Determining the influence of these forces on the genetic make-up of a population is important for understanding evolutionary trajectories and how population structure evolves over time (Funk et al. 2005, Lemay and Russello 2015). For instance, if the amalgamating effect of gene flow is greater than the differentiating effects of genetic drift and divergent selection, then speciation may be inhibited (Lenormand 2002, Bierne et al. 2013). Conversely, if selection is stronger than gene flow, or if strong selection is acting on speciation genes, then divergence can proceed in the presence of gene flow (Feder and Nosil 2010, Ellegren et al. 2012, Feder et al. 2012, Cutter and Payseur 2013).

Because of the antagonistic relationship between natural selection and gene flow, lineages that have broad distributions and a high degree of vagility may experience different speciation processes throughout their geographic range (Proches and Ramdhani 2013a). For example, high dispersal capabilities leading to the colonization of new areas can result in geographically isolated populations where genetic drift and local adaptations play a major role in diversification (Jordan and Snell 2008, McCusker 2010). However, having a high dispersal ability also increases the chance that gene flow can persist among geographically disjunct populations, resulting in secondary contact and possibly impeding the speciation process (Lenormand 2002, Feder et al. 2012, Nosil and Feder 2012, Seehausen et al. 2014).

High dispersal capabilities have resulted in at least 59 lineages of birds having achieved cosmopolitan distributions (Proches and Ramdhani 2013a). Among these lineages, the genus *Anas* (i.e., the dabbling ducks) appears particularly predisposed for global radiations (Johnson and Sorenson 1999, Proches and Ramdhani 2013a, Lavretsky et al. 2014a), which is likely attributable to their vagile tendencies. Within the genus *Anas*, the blue-winged duck complex, named after their prominent blue wing patch, inhabits all continents except Antarctica and comprises six species, including the northern shoveler (*A. clypeata*), the red shoveler (*A. platalea*), the Cape shoveler (*A. smithii*), the Australasian shoveler (*A. rhynchotis*), the blue-winged teal (*A. discors*), and the cinnamon teal (*A. cyanoptera*). Owing to their global distributions and propensity for long distance movements, the blue-winged ducks provide an excellent study system to assess how highly vagile taxa diverge, as species pairs can be found under allopatry, parapatry, and sympatry. Each shoveler species occupies a separate continent (Johnson and Sorenson 1999), where environmental limitations and geographic distance may be causing these taxa to diverge in strict allopatry. Conversely, among the blue/winged cinnamon teal, examples of sympatry, parapatry and allopatry can be observed. For instance, *A. c. septentrionalium* and *A. discors* are sympatric in North America and are both found along the western and northwestern regions of the United States and Canada (Snyder and Lumsden 1951, Swanson et al. 1974, Hubbard 1978, Small 1994, Howell and Webb 1995); this range overlap is relatively recent as *A. discors* has only been documented west of the Rocky Mountains since 1900 (Swanson et al. 1974). *A. discors* has also extended its breeding distribution into the northern regions of South America, where it overlaps the breeding range of *A. c. tropicus (Wilson et al. 2011)*. Moreover, in

South America, cinnamon teal sub-species, *A. c. cyanoptera* and *A. c. orinomus*, are paratric at the extreme edges of their ranges in the Andean highlands of Peru and Argentina (Wilson et al. 2011, Wilson et al. 2013). These parapatric and sympatric distributions suggest that divergence may be proceeding in the presence of gene flow (Wilson et al. 2011).

In this study, I used multilocus data and coalescent methods to examine the evolutionary history of the blue-winged ducks to identify the factors influencing the evolutionary trajectory of this group. Specifically, I reconstructed a species tree using sequences from mitochondrial DNA (mtDNA) and five nuclear DNA (nuDNA) introns, and compared this tree to previously published trees reconstructed from mtDNA (Johnson and Sorenson 1999) and morphology (Livezey 1991). Secondly, I determined the level of population genetic divergence and tested for the presence of gene flow among species pairs diverging in allopatry and sympatry/parapatry. I predicted that taxa diverging in sympatry/parapatry (e.g., *A. c. septentrionalium* and *A. discors*) will show lower levels of divergence and higher rates of gene flow than those diverging in allopatry. However, given the high dispersal capabilities of these species, I predicted that species from different continents will also experience gene flow, especially between the migratory *A. clypeata* and the more sedentary species from the southern hemisphere. In contrast, divergence among allopatric, non-migratory species pairs will be better explained by stochastic processes such as genetic drift. Elucidating the interactions between population connectivity and genetic divergence within a species complex will contribute to the understanding of the dynamics of speciation in globally distributed lineages.

METHODS

Sampling methods and Compiling Genetic Data:

I used previously published genetic datasets of South American cinnamon teal (mtDNA and 5 nuclear introns; Wilson et al. 2013) and North American cinnamon and blue-winged teal (mtDNA and 2 nuclear introns; Wilson et al. 2011). From these studies, vouchered specimens of *A. c. cyanoptera* (*N* = 52 individuals), *A. c. orinomus* (*N* = 49) *A. c. septentrionalium* (*N* = 70), and *A. discors* (*N* = 76) were collected in Argentina, Bolivia, Colombia, Peru, and the United States (Fig. 1-1). Homologous data for northern shovelers from widespread locations in North America (*N* = 48) and Eurasia (*N* = 36) were obtained from Peters et al. (2014). *A. smithii* samples were collected from southern Africa (*N* = 24; Caron et al. 2010; Cumming et al. 2013), vouchered specimens of *A. platalea* were collected from central and southern South America (*N* = 24), and *A. rhynchotis* were collected from Australia (*N* = 19) and New Zealand (*N* = 4) (Fig. 1-1).

DNA sequencing:

Genomic DNA was extracted from muscle tissue or blood using a Qiagen DNeasy tissue kit. For each sample, I amplified and sequenced five nuclear introns (ENO1, PCK1, ODC1, GRIN1, and FGB; collective total of 1,501bp) and the mitochondrial control region (658–983bp), following protocols described by McCracken et al. (2009). PCR products were cleaned using magnetic AMPure beads (Beckman Coulter, INC., Indianapolis, IN). Direct sequencing was performed on an ABI 3730. All sequences for *A. c. cyanoptera*, *A. c. orinomus*, and *A. clypeata*, as well as mtDNA, ENO1 and ODC1

sequences for *A. c. septentrionalium* and *A. discors*, have been published previously (McCracken et al. 2009, Wilson et al. 2011, Wilson et al. 2013, Peters et al. 2014).

To determine the gametic phase for each intron, I used PHASE version 2.1.1 (Stephens et al. 2001). The program PHASE is a Bayesian statistical method for reconstructing haplotypes form genetic data while incorporating Hardy-Weinberg equilibrium and linkage disequilibrium. I constructed input files using seqPHASE (Flot 2010), treating the blue-winged/cinnamon teal and the shoveler lineages in separate analyses. For each analysis, I ran 1,000 iterations with a burn-in of 1,000 generations.

Population Structure and Genetic Diversity:

To estimate levels of genetic differentiation (F_{ST}: the proportion of total genetic diversity partitioned between populations) and nucleotide diversity (π: the average proportion of nucleotide differences for a given locus within a population) for each locus, I used DNAsp version 5.10 (Librado and Rozas 2009). I also used the Bayesian clustering method in the program SplitsTree version 4.13 that provides a framework for evolutionary analysis by combining both phylogenetic trees and networks to visualize genetic variation within and among populations (Ayling and Brown 2008). I constructed splits trees using consensus sequences of the concatenated nuDNA (with heterozygous sites coded with IUPAC ambiguity codes) and a second tree using mtDNA only. I used P-distances with ambiguities treated as the average state (e.g., T vs. C is a distance of 1.0, whereas Y vs. C is a distance of 0.5).

Coalescent Analysis

I estimated phylogenetic relationships within the blue-winged complex using the coalescent program *BEAST version 1.7.4 (Heled and Drummond 2010). *BEAST uses Bayesian analysis incorporating a Markov Chain Monte Carlo (MCMC) in species-tree estimation (Heled and Drummond 2010). Two separate phylogenetic analyses were performed using the constant population size prior; one analysis included the five nuclear loci and mitochondrial DNA and the other included only nuclear loci. Using MEGA v6.0 (Humer et al. 1994) I estimated the optimal substitution rate based on the maximum likelihood of different models. The HKY substitution model, which allows for different substitution rates between transitions and transversions, was selected for all loci. Each phylogeny included the same five outgroup species, including the Hottentot teal (*A. hottentota*), Puna teal (*A. puna*), garganey (*A. querquedula*), silver teal (*A. versicolor*), and Baikal teal (*A. formosa*) (McCracken et al. 2009; unpubl. data), and I treated populations of *A. clypeata* and *A. rhynchotis* from different landmasses as separate operational taxonomic units (i.e., North America vs. Eurasia and Australia vs. New Zealand, respectively). For each analysis, I ran 10,000,000 iterations, sampling every 1,000 MCMC steps following a burn-in of 1,000,000 steps. The posterior set of trees was analyzed in DensiTree (Bouckaert & Heled 2014) to create a phylogeny where MCMC simulations are visible and not averaged together. To obtain a consensus tree with posterior distributions, I used TreeAnnotator version 1.7 (Drummond et al. 2012) discarding the first 1,000 trees as burn-in.

To estimate levels of gene flow, I used the Isolation with Migration model IMa2 (Hey and Nielsen 2004). The IMa2 model is an evolutionary model that uses genetic data to infer evolutionary histories and different demographic parameters (Hey and Nielsen

2004). Because the full data set was too cumbersome to analyze using IMa2, I ran four separate analyses: (1) a four-population model that included the four species of shovelers (allopatric populations), (2) a four-population model that included the blue-winged teal and the three sub-species of cinnamon teals (allopatric and parapatric populations), (3) a three-population model that included *A. clypeata* from North America, *A. discors*, and the North American cinnamon teal *A. c. septentrionalium* (parapatric and sympatric populations), and (4) a two-population model that included *A. platalea* and lowland cinnamon teal, *A. c. cyanoptera*, from South America (sympatric populations). Here, I estimated three different evolutionary parameters (a total of 22 parameters for four-population models, 13 parameters for the three-population model, and 6 parameters for the two-population model). These parameters included Θ (where $\Theta = 4N_e\mu$, and N_e is the effective population size and μ is the mutation rate per locus per generation) for each contemporary population and the ancestral populations, $2Nm$ (which is $\Theta m_{im}/2$, and m_{im} is the migration rate relative to the mutation rate estimated in IMa2, N is the population size, and m is the proportion of the population consisting of immigrants per generation) between each pair of species/sub-species, and t (where $t = T/\mu$, and T is the number of years since divergence). Because IMa2 assumes only recombination between loci and not within a locus, I used IMgc (Woerner et al. 2007) to detect violations of the four-gamete test and to truncate loci to be consistent with no recombination. I defined the species trees for each run based on the results of *BEAST (see above). I ran one cold chain and 59 hot chains with geometric heating for 1,000,000 generations as a burn in and then sampled parameters every 200 generations for 30,000,000 iterations. The analysis was replicated

with a different random number seed to check for convergence, and all ESS values were >50 in all runs.

RESULTS

Genetic structure of the blue-winged ducks

Among nuclear loci, nucleotide diversity ranged between 0 and 0.022 substitutions/site with a mean of 0.0072 substitutions/site. For mtDNA, nucleotide diversity ranged from 0.0001 to 0.0096 substitutions/site with a mean of 0.0044 substitutions/site. Across all pairwise comparisons, F_{ST} ranged from 0.080 to 0.938 for all nuclear loci, and from 0.072 to 0.955 for mtDNA (Table 1-1). Overall differentiation was highest between the shovelers and the blue-wing/cinnamon teal group and among all pairwise comparisons of shoveler except *A. clypeata* and *A. rhynchotis* (pairwise F_{ST} > 0.2 in most cases; Table 1-1). Neighbor-net trees for nuDNA showed that individuals mostly clustered with other individuals from the same sub-lineage; the blue-winged/cinnamon teal sub-lineage and shoveler sub-lineage (Fig. 1-2a). Within the shoveler sub-lineage, *A. platalea* and *A. smithii* appeared to be the most differentiated; all individuals from each of these species cluster with other individuals of the same species to the exclusion of other species. These patterns of clustering coincide well with pairwise F_{ST} values. However, *A. clypeata* and *A. rhynchotis* had weaker levels of genetic differentiation and clustering analysis showed a high level of admixture between these species (Fig. 1-2a). Additionally, the blue-winged/cinnamon teal sub-lineage had very weak clustering which is also supported by low levels of genetic differentiation (Table 1-1). Interestingly, cases of inter-lineage clustering were present where four samples of *A.*

discors and one sample of *A. c. septentrionalium* clustered with the shovelers, and one sample of *A. rhynchotis* clustered within the blue-winged/cinnamon teals (Fig. 1-2a).

In contrast to nuDNA, the mtDNA neighbor-net tree is consistent with high genetic differentiation within the shoveler sub-lineage, with most individuals grouping within species-specific lineages (mean F_{ST} = 0.826). The only two exceptions are that a mtDNA haplotype from one *A. clypeata* individual fell outside the *A. clypeata*/*rhynchotis* cluster, and one *A. rhynchotis* mtDNA haplotype (sampled from four individuals from New Zealand) clustered within the *A. clypeata* cluster (Fig. 1-2b). Much lower differentiation was observed within the blue-winged/cinnamon teal sub-lineage, although haplotypes from *A. c. septentrionalium* mostly clustered within a distinct group (also see Wilson et al. 2011).

Phylogenetic Reconstruction

The analysis of nuDNA in *BEAST recovered the shoveler sub-lineage and the blue-winged/cinnamon sub-lineage to be reciprocally monophyletic (Fig. 1-3). This relationship was also supported by levels of genetic differentiation for both nuDNA and mtDNA (F_{ST} = 0.254 and 0.860 between sub-lineages, respectively). Shallow divergences were recovered between *A. c. septentrionalium* and *A. discors*, between *A. c. cyanoptera* and *A. c. orinomus* (posterior support = 1.0), between Old and New World *A. clypeata* (posterior support = 0.99) and between Australasian sub-species, *A. r. rhynchotis* and *A. r. variegata* (posterior support = 0.99). I found weak posterior support for *A. smithii* as sister to the *A. clypeata*/*rhynchotis* group (posterior support = 0.37), and for the monophyly of the shovelers (posterior support = 0.47). Note that an appreciable number of posterior trees recovered *A. platalea* as sister to the blue-winged/cinnamon teal group

(Fig. 1-3), rendering the shovelers as paraphyletic. When I included both mtDNA and nuDNA, the *BEAST analysis did not change the tree topology, and most relationships received higher posterior support, including the monophyly of the shoveler sub-lineage (posterior support = 0.48).

Estimates of gene flow (2*Nm*)

Within the shoveler sub-lineage, I could not reject a scenario of no gene flow into *A. clypeata*, *A. smithii*, or *A. platalea* from any of the other shoveler species (Fig. 1-4a). In general, posterior distributions peaked at or near zero and confidence intervals were narrow. In contrast, I found evidence of low, asymmetrical gene flow into *A. rhynchotis* from both *A. clypeata* (2*Nm* = 0.5 migrants per generation) and *A. smithii* (2*Nm* = 0.16 migrants per generation) (Fig. 1-4c). In both cases, however, the 95% highest posterior distribution included the lowest bin of migration rates, suggesting that the data are also consistent with complete isolation.

In contrast, within the blue-winged/cinnamon teal sub-lineage, prominent traces of gene flow were detected from *A. c. orinomus* into *A. c. cyanoptera* (2*Nm* = six migrants per generation) and in the reverse direction (2*Nm* = two migrants per generation) (Fig. 1-5a,b). Strong levels of intra-continental gene flow were also found from *A. discors* into *A. c. septentrionalium* (2*Nm* = nine migrants per generation) and in the opposite direction from *A. c. septentrionalium* into *A. discors* (2*Nm* = five migrants per generation). However, confidence intervals overlapped zero in the latter direction. Although I found some evidence of intercontinental gene flow from *A. discors* into *A. c. cyanoptera* (2*Nm* = 0.81 migrants per generation, but note the high posterior probability

associated with the lowest bin in Fig. 1-5a), my overall results suggested more extensive gene flow between parapatric populations than between allopatric populations. Finally, I did not find evidence of gene flow between sympatric species of shovelers and *A. discors* or cinnamon teal in either North America or South America; all posterior distributions of 2*Nm* peaked at or near zero (Fig. 1-6).

DISCUSSION

Phylogenetic relationships

In this study I used a multi-locus approach to examine the processes shaping patterns of genetic divergence within the blue-winged duck complex. My multi-locus dataset supported reciprocal monophyly between the shovelers and blue-winged/cinnamon teals, which agrees with a morphological tree reconstructed from 157 phenotypic characters (Livezey 1991). In contrast, a mtDNA gene tree recovered the shovelers as being paraphyletic with respect to the blue-winged/cinnamon teals (Johnson and Sorenson 1999). Specifically, *A. platalea* was sister to a clade comprising the remaining shovelers, blue-winged teal, and cinnamon teals. However, this sister relationship received low (<50%) bootstrap support, and the many reticulations in the neighbor-net tree further illustrates the uncertainty in the mtDNA gene tree (Fig. 1-2a,b). In my multi-locus species tree reconstructions, I found a small set of posterior trees supporting the paraphyly of the shovelers, but in this case, *A. platalea* was sister to the blue-winged/cinnamon teals (Fig. 1-3). Although a monophyletic shoveler lineage was the best-supported species tree when analyzing nuDNA, this lineage also received low

posterior support (<50%), and therefore, additional data may be needed to fully resolve phylogenetic relationships within this group of ducks.

Assuming that the mtDNA gene tree accurately reflects the history of this marker and that my species tree accurately reflects the history of population divergence and speciation, my results suggest conflict between mitochondrial and nuclear markers. Mito-nuclear discord can result from a number of processes, including stochastic lineage sorting, hybridization, and sex-biased gene flow (Toews and Brelsford 2012). For example, because populations of mitochondrial loci are a quarter the size of nuDNA, divergence and lineage sorting can happen at a much faster rate for mtDNA, causing evolutionary discords between the two genomes. The short internodes at the base of the mtDNA and nuDNA phylogenies (Johnson and Sorenson, 1999; Figs. 1-2 & 1-3) suggest rapid diversification, and therefore, I can expect high levels of among-locus conflict being driven by stochastic lineage sorting. As found in other studies (Machado et al. 2002, Faure et al. 2009, Bulgarella et al. 2012, Paupério et al. 2012, Lavretsky et al. 2014b), using a multi-locus approach allows us to account for stochastic lineage sorting and conflict among loci that might otherwise lead to biases when inferring evolutionary relationships based on a single marker type. However, this conflict likely contributes to low support for phylogenetic relationships in the absence of large datasets.

Despite the conflict in relationships between shovelers and blue-winged/cinnamon teals recovered from mitochondrial and nuclear markers, my species tree agrees with the mitochondrial genealogy to some extent (Johnson and Sorenson 1999). For example, both mtDNA and the nuDNA species tree support a sister relationship between *A. smithii* and the *A. clypeata*/*rhynchotis* group, whereas morphological characteristics (including

skeletal and plumage characteristics) suggest that *A. clypeata*/*rhynchotis* is sister to *A. platalea* (Livezey 1991). Patterns of dichromatism are prevalent in *A. platalea*, *A. clypeata*, and *A. rhynchotis* but not in *A. smithii*, which might have contributed to this morphological grouping. In addition, my results agree with the mtDNA gene tree in suggesting that cinnamon teals are paraphyletic with respect to *A. discors*, although in that case *A. discors* was sister to *A. c. cyanoptera* rather than *A. c. septentrionalium* (Johnson and Sorenson 1999, Wilson *et al.* 2011). Wilson *et al.* (2011) suggested a South American origin for the blue-winged/cinnamon teal complex; specifically, a South American cinnamon teal ancestor might have given rise to *A. discors* and *A. c. septentrionalium*, although the number of colonization and the order of divergence events were unambiguous (Wilson et al. 2011). The paraphyly of *A. cyanoptera* with respect to *A. discors* in my multi-locus species tree supports a single colonization of North America, followed by divergence between *A. discors* and *A. c. septentrionalium*. However, estimating species trees in *BEAST requires an assumption of no gene flow between operational taxonomic units (Heled and Drummond 2010), but I found strong evidence of gene flow between *A. discors* and *A. c. septentrionalium* (see below). Gene flow, in this case, could have resulted from secondary contact following two independent colonization events of North America, which could mislead phylogenetic reconstructions. Analyzing larger datasets with methods that jointly estimate migration and phylogenies may be necessary to conclusively resolve the placement of *A. discors* within the blue-winged/cinnamon teal complex.

Estimated number of migrants per generation (2Nm)

The high dispersal abilities of the blue-winged duck complex have led to bouts of gene flow within both sub-lineages. Specifically, within the shoveler sub-lineage, evidence of unidirectional gene flow was found among allopatric species from different continents (from *A. clypeata* into *A. rhynchotis* and from *A. smithii* into *A. rhynchotis*), although confidence intervals included zero for both comparisons (Fig. 1-5a). Although gene flow is unexpected from *A. smithii*, the wintering range of the migratory *A. clypeata* has expanded into the Philippines and Malaysia (Van Weerd and Van der Ploeg 2004), and vagrant individuals from Australia and New Zealand were reviewed in Close and Jaensch (1981) and Marchant and Higgins (1990), as well as reported in public databases, such as eBird (ebird.org; Sullivan et al. 2009) and the Atlas of Australian Birds (birddats.com.au)). These vagrants provide opportunities for hybridization and gene flow between these otherwise allopatric species. For hybridization to occur it is likely that a small fraction of individuals must take up at least semi-residency in either Australia or New Zealand during the breeding season. As sightings in the Philippines have been reported between May and August (Van Weerd and Van der Ploeg 2004), it appears that at least some vagrant *A. clypeata* males have at least partially suspended their migration back to the typical northern breeding grounds. Although females of these two species are difficult to differentiate and therefore it is unknown if females are present year-round as well, the year-round presence of males raises the possibility they could be reproductively active thus providing an avenue for potential hybridization. Interestingly, all four samples of *A. rhynchotis* from New Zealand had mtDNA haplotypes that clustered more closely with haplotypes from *A. clypeata* than with *A. rhynchotis* from Australia suggesting

mtDNA introgression and past female-mediated gene flow. To my knowledge there are no confirmed wild hybrids, suggesting that hybridization is probably rare.

Congruent with previous research (Wilson *et al.* 2011), I found evidence of gene flow between parapatric taxa in South America (*A. c.cyanoptera* and *A. c. orinomus*) and North America (*A. c. septentrionalium* and *A. discors*) within the blue-winged/cinnamon teal sub-lineage. Within central South America, cinnamon teal subspecies occur in sympatry at the extreme edges of the distribution of *A. c. orinomus* within the Central High Andes (Wilson et al. 2011, 2013). Although sightings of lowland taxa in the highlands, and vice versa, have been reported for many species, the hypoxic environment of the High Andes likely restricts gene flow into the highlands as seen in other waterfowl (Bulgarella *et al.* 2012,McCracken et al. 2009a, McCracken et al. 2009b). Intermixing between subspecies also could occur at mid-elevations as observed in the crested duck (*Lophonetta specularioides*), which is characterized by a well-defined hybrid zone at an elevation of ~2000 m (Bulgarella *et al.* 2012); however, cinnamon teal appear to be rare at mid-elevation as the lowest record for *A. c. orinomus* is approximately 1500 meters based on a single bird (Wilson et al. 2013).

Within North America, episodes of gene flow may be dictated by seasonal breeding migrations between *A. c. septentrionalium* and *A. discors*. Distributions of *A. c. septentrionalium* and *A. discors* overlap more extensively in the western region of the United States and Canada where hybridization does occur (Lokemoen et al. 1990, Gammonley and Fredrickson 1995, Howell and Webb 1995, Wilson et al. 2011), although the extent of hybridization is unknown. I detected non-zero gene flow (CIs do not overlap zero) indicating that these two taxa maintained some level of genetic

connectivity after divergence. This level of genetic exchange and directionality could potentially reflect the recent expansion of *A. discors* breeding ranges, although more data are needed to test this hypothesis.

Interestingly, Wilson et al. (2011) detected non-zero values for migration rates between *A. c. cyanoptera* and *A. c. sepetentrionalium*, suggesting inter-continental gene flow within the sub-lineage. However, this was not the case in this study where only intra-continental gene flow was detected as all confidence intervals between *A. c. cyanoptera* and *A. septentrionalium* broadly overlapped zero. One reason for this could reside in resolution power within genetic markers. For example, Wilson et al. (2011) used a total of three loci (two nuclear loci and one mitochondrial locus), which might have resulted in a lack of resolution for coalescent analyses (Cruickshank and Hahn 2014). By using six total loci, I likely increased the power and resolution. An alternative contributing factor could be Wilson et al.'s (2011) use of pairwise comparisons in the two-population IM models. Specifically, one of the assumptions when using IM is that there is no gene flow among populations that are not included in the analysis (Hey and Nielsen 2004). Thus, when running only a two-population model, any shared genetic variation that cannot be explained by ancestral variation between the two taxa may give a false positive of non-zero migration rates. In contrast, running analyses that include all taxa incorporates a full-migration matrix, accounting for sources of variation from all other taxa, and potentially reducing the occurrence of false positives. Overall, estimates of gene flow within the blue-winged/cinnamon teal suggest that geographic distance and other environmental variables between North and South America likely create strong barriers to gene flow resulting in allopatric divergence and speciation.

CONCLUSION

Although speciation varies among lineages (Bierne et al. 2013, Proches and Ramdhani 2013b), studying the evolutionary histories among closely related taxa, at different stages of the speciation continuum, can reveal the propensity of different evolutionary forces to act on the process of divergence and ultimately speciation (Johnson and Sorenson 1999, Lenormand 2002). Here I show that the evolutionary relationships within the blue-winged duck complex suggests two reciprocally monophyletic sub-lineages. I suggest that gene flow is most common among parapatric sister taxa (especially in the case of *A. c. septentrionalium* and *A. discors*) and that more distantly related taxa experience nearly complete isolation and may be diverging via allopatric speciation. In the latter case, larger geographical features (e.g., mountain ranges, bodies of water, and geographical distance) may be playing a role in reducing long distance dispersal events and subsequently decreasing the probability of hybridization. However, the high dispersal capability of the migratory *A. clypeata* may facilitate gene flow among otherwise allopatric populations. Because the blue-winged ducks show diverse patterns of speciation, this complex can serve as a model system for identifying how the genome diverges under different evolutionary processes and how genetic variation is partitioned among highly dispersive taxa.

REFERENCES

Bierne, N., Gagnaire, P. and David, P. 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation - Curr. Zool 59: 72-86.

Bulgarella, M., Peters, J. L., Kopuchian, C., Valqui, T., Wilson, R. E. and McCracken, K. G. 2012. Multilocus coalescent analysis of haemoglobin differentiation between low-and high-altitude populations of crested ducks (*Lophonetta specularioides*) - Mol. Ecol. 21: 350-368.

Close, D.H. and Jaensch, R.P. 1981. Northern Shoveler at Coongie Lake. S. Austral. Ornithol. 28: 178-179.

Cruickshank, T. E. and Hahn, M. W. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. - Mol. Ecol. 23: 3133-57.

Cutter, A. D. and Payseur, B. A. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species - Nat. Rev. Genet. 14: 262-274.

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S. and Wolf, J. B. W. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers - Nature 491: 756-760.

Faure, B., Jollivet, D., Tanguy, A., Bonhomme, F. and Bierne, N. 2009. Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels - PLoS One 4: e6485.

Feder, J. L., Egan, S. P. and Nosil, P. 2012. The genomics of speciation-with-gene-flow - Trend Genet. 28: 342-350.

Feder, J. L. and Nosil, P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation - Evolution 64: 1729-1747.

Flot, J. 2010. SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments - Mol. Ecol. Resour. 10: 162-166.

Funk, W. C., Blouin, M. S., Corn, P. S., Maxell, B. A., Pilliod, D. S., Amish, S. and Allendorf, F. W. 2005. Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape - Mol. Ecol. 14: 483-496.

Gammonley, J. H. and Fredrickson, L. H. 1995. Life history and management of the blue-winged teal. - In: Cross, D. H. and Vohs, P. (eds.), Waterfowl Management Handbook. United States Department of the Interior, pp. 13.1.8.

Heled, J. and Drummond, A. J. 2010. Bayesian inference of species trees from multilocus data - Mol. Biol. Evol. 27: 570-580.

Howard, P. J. 1996. A New Zealand record of the northern shoveler. Notornis 15:253

Howell, S. N. and Webb, S. 1995. A guide to the birds of Mexico and northern Central America. - Oxford University Press, Oxford.

Hubbard, J. P. 1978. Revised check-list of the birds of New Mexico. - New Mexico Ornithological Society.

Johnson, K. P. and Sorenson, M. D. 1999. Phylogeny and biogeography of dabbling ducks (genus: *Anas*): a comparison of molecular and morphological evidence - Auk 116: 792-805.

Jordan, M. A. and Snell, H. L. 2008. Historical fragmentation of islands and genetic drift in populations of Galapagos lava lizards (*Microlophus albemarlensis* complex) - Mol. Ecol. 17: 1224-1237.

Lavretsky, P., Hernández-Baños, B. E. and Peters, J. L. 2014a. Rapid radiation and hybridization contribute to weak differentiation and hinder phylogenetic inferences in the New World Mallard complex (*Anas* spp.) - Auk 131: 524-538.

Lavretsky, P., McCracken, K. G. and Peters, J. L. 2014b. Phylogenetics of a recent radiation in the Mallards and allies (Aves: *Anas*): Inferences from a genomic transect and the multispecies coalescent - Mol. Phylogenet. Evol. 70: 402-411.

Lemay, M. A. and Russello, M. A. 2015. Genetic evidence for ecological divergence in kokanee salmon - Mol. Ecol. 24: 798-811.

Lenormand, T. 2002. Gene flow and the limits to natural selection - Trends Ecol. Evol. 17: 183-189.

Livezey, B. C. 1991. A phylogenetic analysis and classification of recent dabbling ducks (Tribe Anatini) based on comparative morphology - Auk 108: 471-507.

Lokemoen, J. T., Duebbert, H. F. and Sharp, D. E. 1990. Homing and reproductive habits of mallards, gadwalls, and blue-winged teal - Wildl. Monogr. 106: 3-28.

Machado, C. A., Kliman, R. M., Markert, J. A. and Hey, J. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives - Mol. Biol. Evol. 19: 472-488.

Marchant, S. and Higgins, P.J. 1990. Handbook of Australian, New Zealand and Antarctic Birds. Volume 1. Part B. Australian Pelican to Ducks. Melbourne: Oxford University Press.

McCracken, K. G., Barger, C. P., Bulgarella, M., Johnson, K. P., Kuhner, M. K., Moore, A. V., Peters, J. L., Trucco, J., Valqui, T. H. and Winker, K. 2009a. Signatures of high- altitude adaptation in the major hemoglobin of five species of andean dabbling ducks - Am. Nat. 174: 631-650.

McCracken, K. G., Bulgarella, M., Johnson, K. P., Kuhner, M. K., Trucco, J., Valqui, T. H., Wilson, R. E. and Peters, J. L. 2009b. Gene flow in the face of countervailing selection: adaptation to high-altitude hypoxia in the betaA hemoglobin subunit of yellow-billed pintails in the Andes - Mol. Biol. Evol. 26: 815-827.

McCusker, M. R. and Bentzen, P. 2010. Historical influences dominate the population genetic structure of a sedentary marine fish, Atlantic wolffish (*Anarhichas lupus*), across the North Atlantic Ocean - Mol. Ecol. 19: 4228-4241.

Nosil, P. and Feder, J. L. 2012. Genomic divergence during speciation: causes and consequences - Philos. Trans. R. Soc. Lond. B. Biol. Sci 367: 332-342.

Paupério, J., Herman, J., Melo- Ferreira, J., Jaarola, M., Alves, P. and Searle, J. 2012. Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia - Mol. Ecol. 21: 6015-6032.

Peters, J. L., Winker, K., Millam, K. C., Lavretsky, P., Kulikova, I., Wilson, R. E., Zhuravlev, Y. N. and McCracken, K. G. 2014. Mito-nuclear discord in six congeneric lineages of Holarctic ducks (genus *Anas*) - Mol. Ecol. 23: 2961-2974.

Proches, S. and Ramdhani, S. 2013a. Eighty-three lineages that took over the world: a first review of terrestrial cosmopolitan tetrapods - J. Biogeogr. 40: 1819-1831.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G., Bank, C., Braennstroem, A., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., Lindholm, A. K., Lucek, K., Maan, M. E., Marques, D. A., Martin, S. H., Matthews, B., Meier, J. I., Moest, M., Nachman, M. W., Nonaka, E., Rennison, D. J., Schwarzer, J., Watson, E. T., Westram, A. M. and Widmer, A. 2014. Genomics and the origin of species - Nat. Rev. Genet. 15: 176-192.

Small, A. 1994. California birds. Their status and distribution. - IBIS Press, Vista, CA.

Snyder, L. L. and Lumsden, H. G. 1951. Variation in *Anas cyanoptera* - Occ. Pap. Roy. Ontario Mus. Zool. 10: 1-18.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D. and Kelling, S. 2009. eBird: A citizen-based bird observation network in the biological sciences - Biol. Conserv. 142: 2282-2292.

Swanson, G. A., Meyer, M. I. and Serie, J. R. 1974. Feeding ecology of breeding blue-winged teals - J. Wildl. Manage. 38: 396-407.

Toews, D. P. and Brelsford, A. 2012. The biogeography of mitochondrial and nuclear discordance in animals - Mol. Ecol. 21: 3907-3930.

Van Weerd, M. and Van der Ploeg, J. 2004. Surveys of wetlands and waterbirds in Cagayan valley, Luzon, Philippines - Forktail 20: 33-39.

Wilson, R. E., Eaton, M. D., Sonsthagen, S. A., Peters, J. L., Johnson, K. P., Simarra, B. and McCracken, K. G. 2011. Speciation, subspecies divergence, and paraphyly in the cinnamon teal and blue-winged teal - Condor 113: 747-761.

Wilson, R. E., Peters, J. L. and McCracken, K. G. 2013. Genetic and phenotypic divergence between low-and-high-altitude populations of two recently diverged cinnamon teal subspecies - Evolution 67: 170-184.

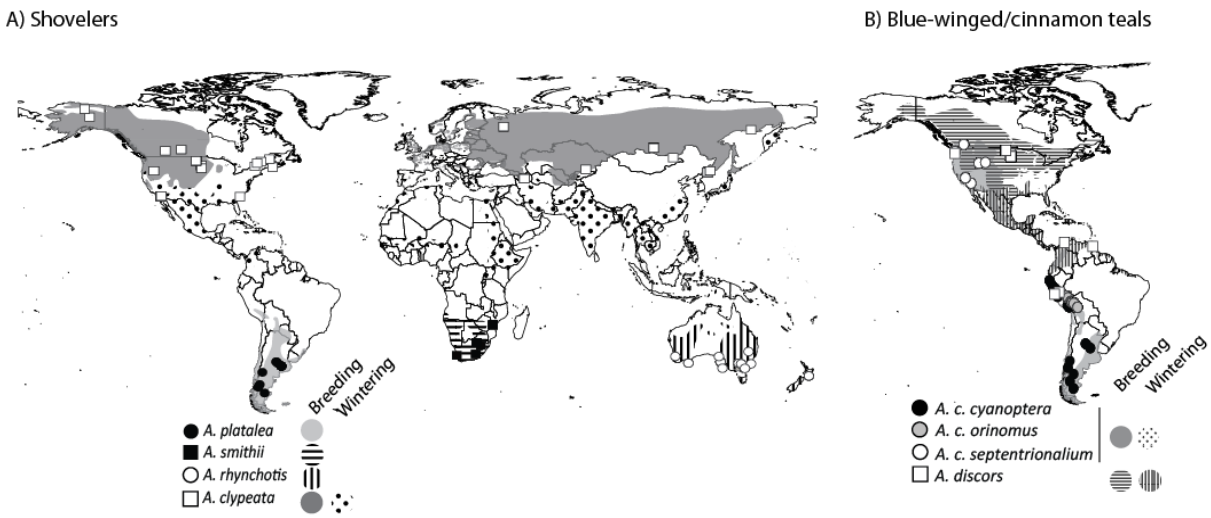

Figure 1-1. The distributions (shaded regions) and sampling locations (symbols) for A) four species of shovelers and B) two species and three sub-species from the blue-winged/cinnamon teal.

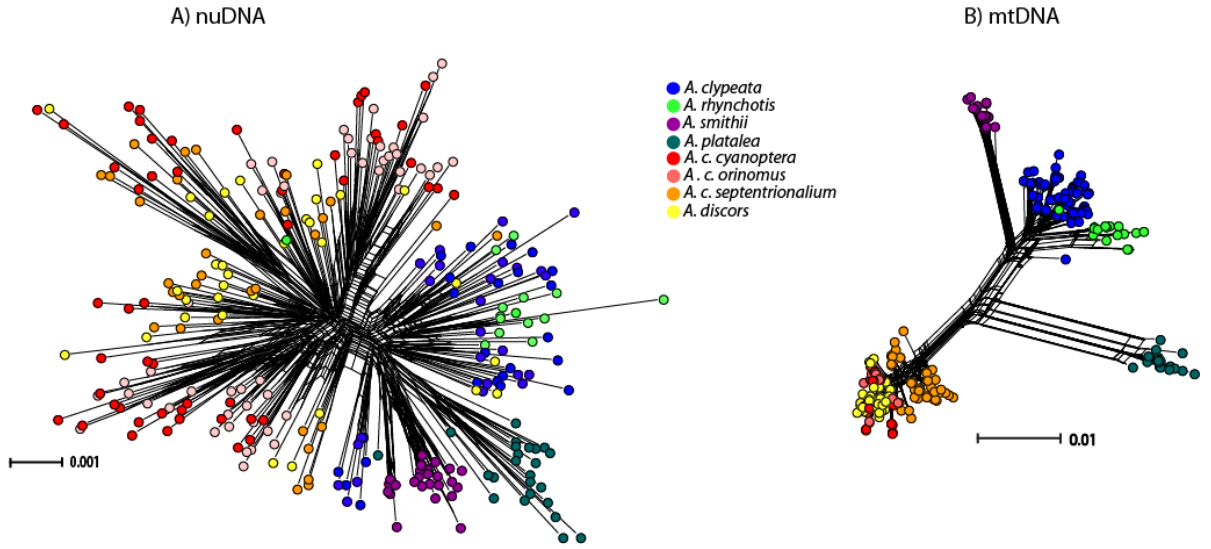

Figure 1-2. Neighbor-net trees of the blue-winged duck complex for concatenated nuDNA (A) at five loci (1,728 aligned nucleotides) and mtDNA (B) (652 aligned nucleotides). Note that individuals representing the shoveler sub-lineage are displayed in "cold" colors (i.e., blue, purple, teal, and green), and individuals representing the cinnamon teal sub-lineage are displayed in "hot" colors (i.e., red, orange, yellow, and pink).

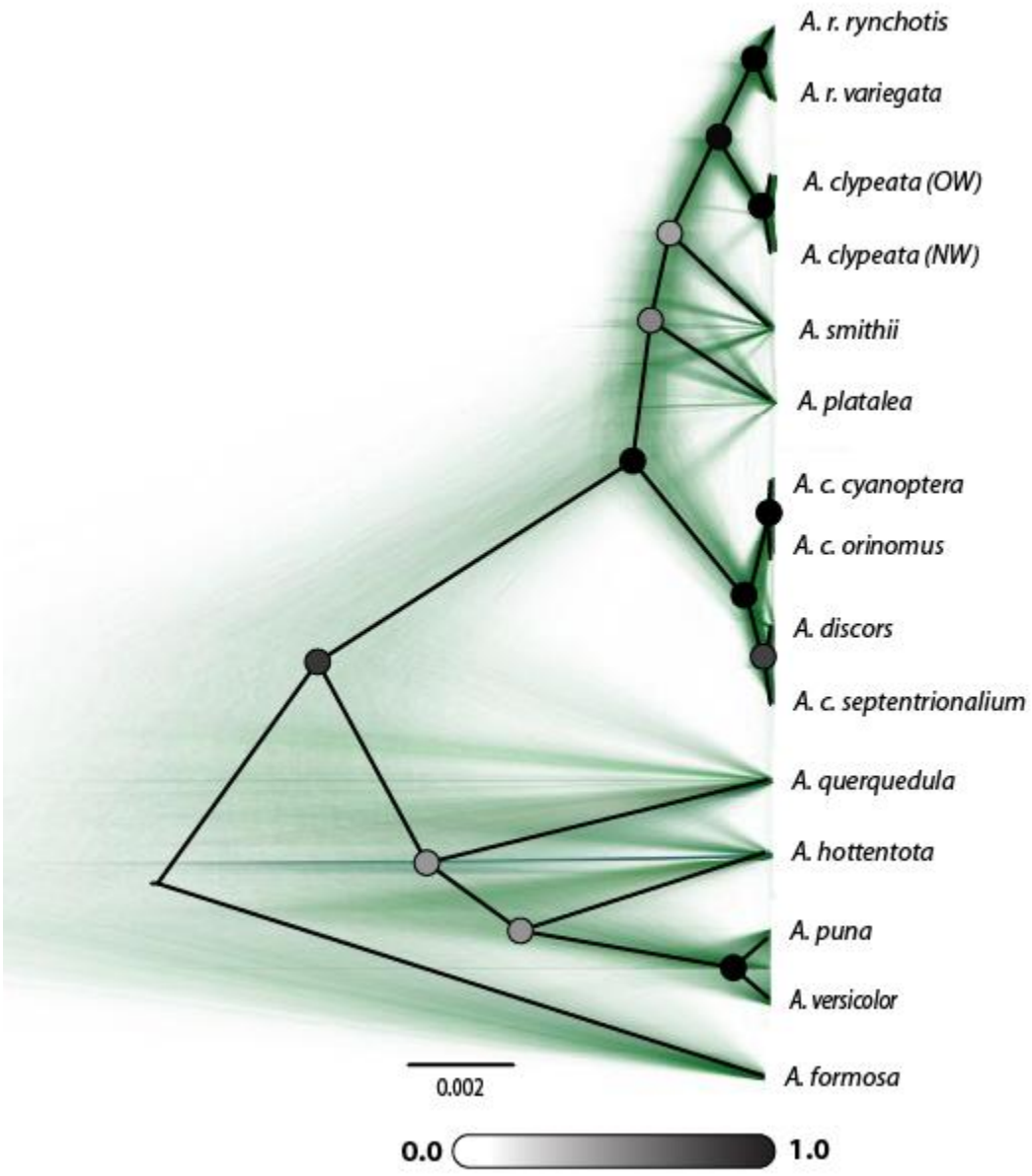

Figure 1-3. The evolutionary relationships of the blue-winged duck complex reconstructed using *BEAST and five nuclear introns. Shaded regions illustrate all 1,000 posterior trees and the dark lines show the consensus phylogeny. The shaded circles represent the overall posterior support for each node, where lighter colors reflect lower posterior support and darker colors reflect stronger posterior support as illustrated in the lower bar.

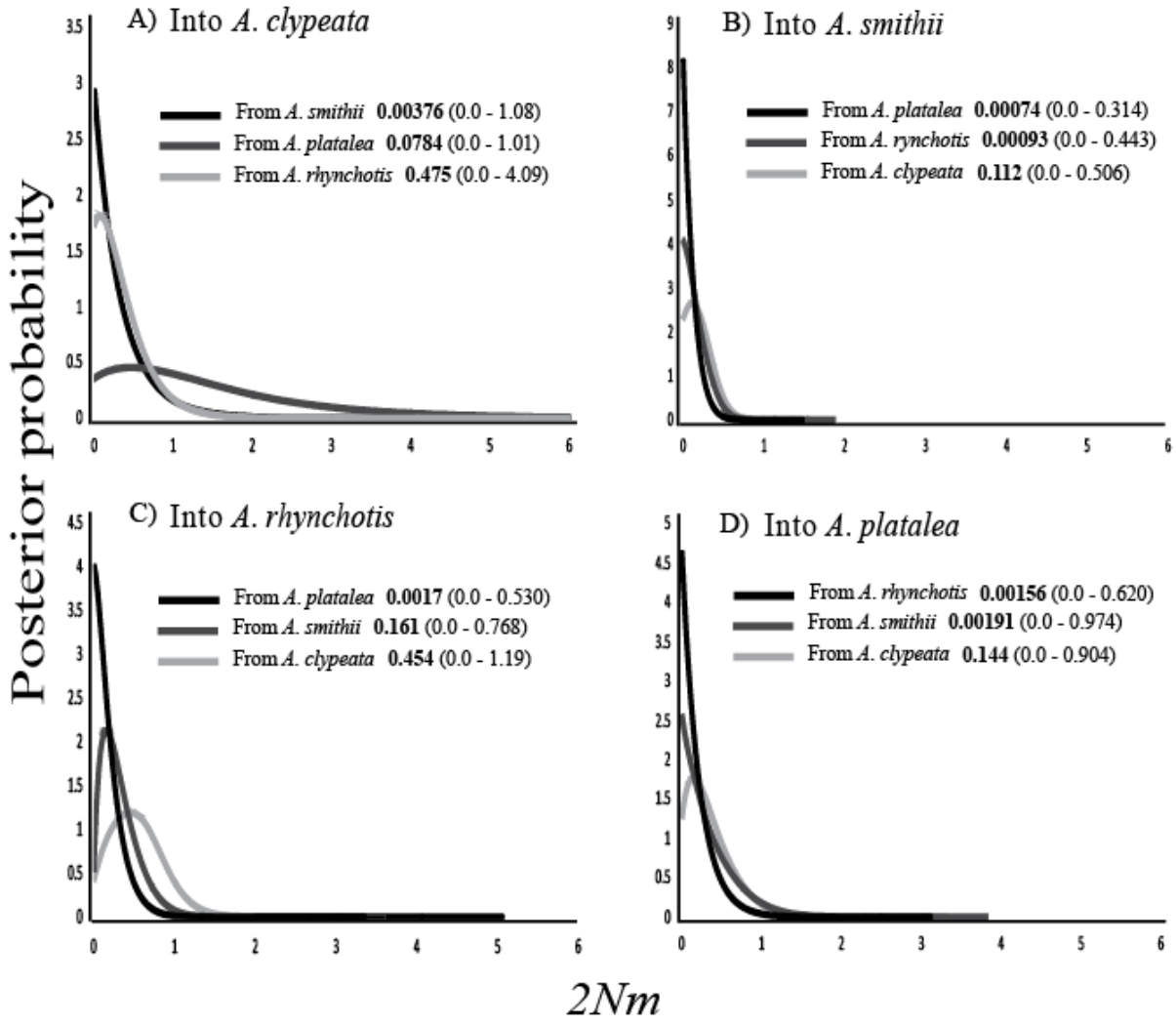

Figure 1-4. Estimates of the number of immigrants per generation for each of the shoveler sub-lineages: *A. clypeata* (A), *A. smithii* (B), *A. rhynchotis* (C), and *A. platalea* (D). Values in bold represent the highest posterior probability and the values in parentheses are the 95% highest posterior densities.

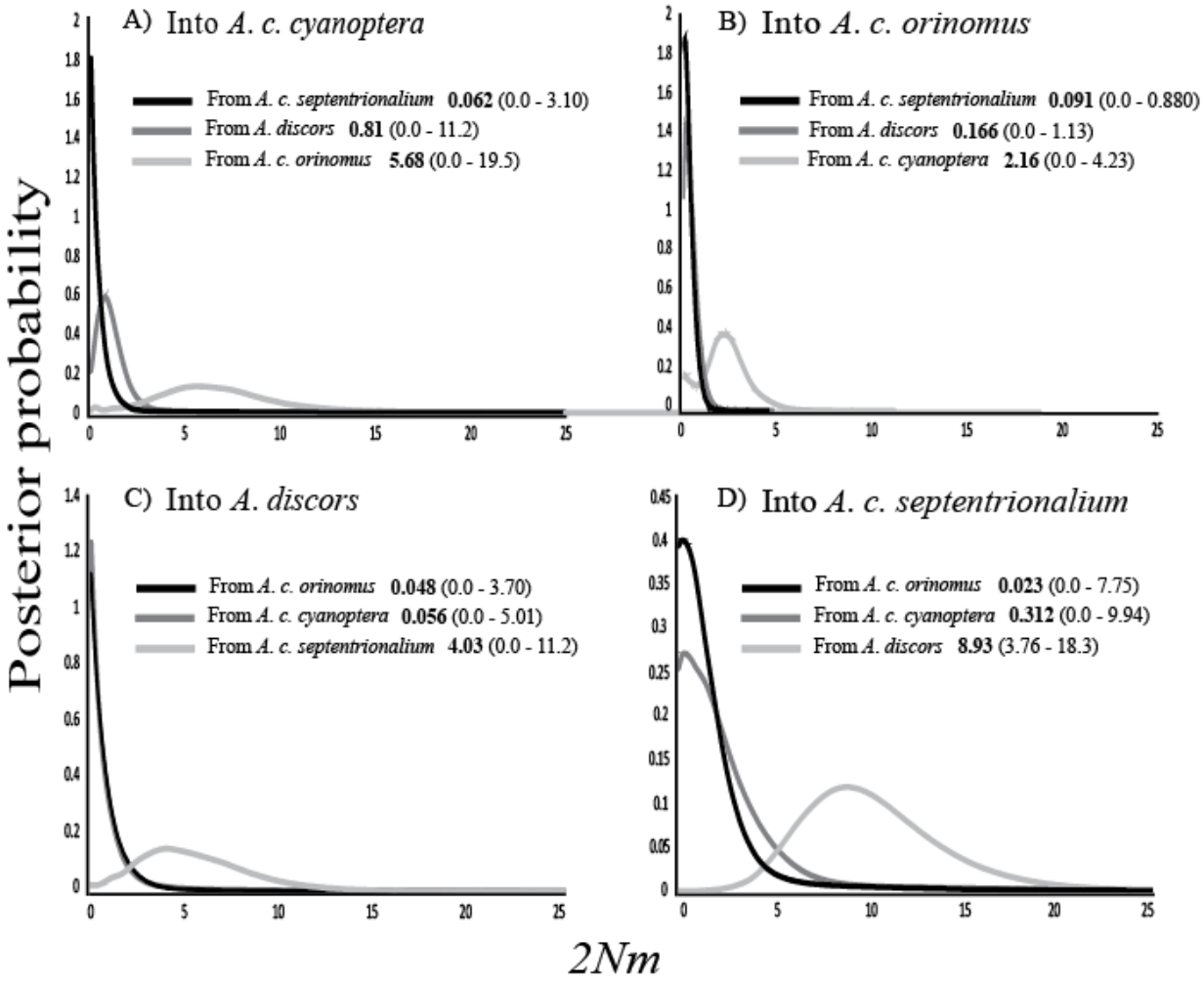

Figure 1-5. Estimates of the number of immigrants per generation for each taxon of the blue-winged/cinnamon teal complex: *A. c. cyanoptera* (A), *A. c. orinomus* (B), *A. discors* (C), and *A. c. septentrionalium* (D). Values in bold represent the highest posterior probability and the values in parentheses are the 95% highest posterior densities.

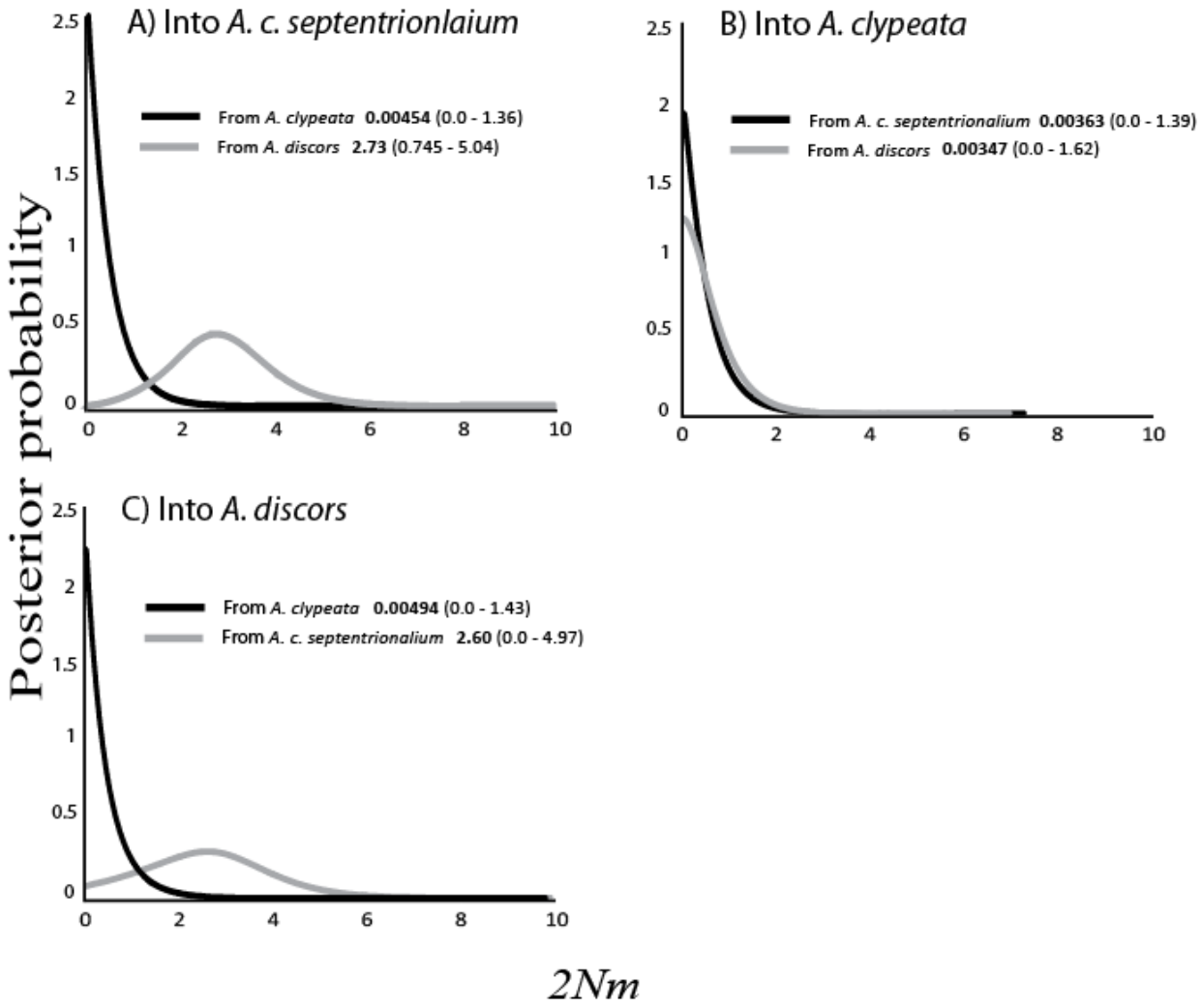

Figure 1-6. Estimates of the number of immigrants per generation for each North American comparison: *A. c. septentrionalium* (A), *A. clypeata* (B), and *A. discors* (C). Values in bold represent the highest posterior probability and the values in parentheses are the 95% highest posterior densities.

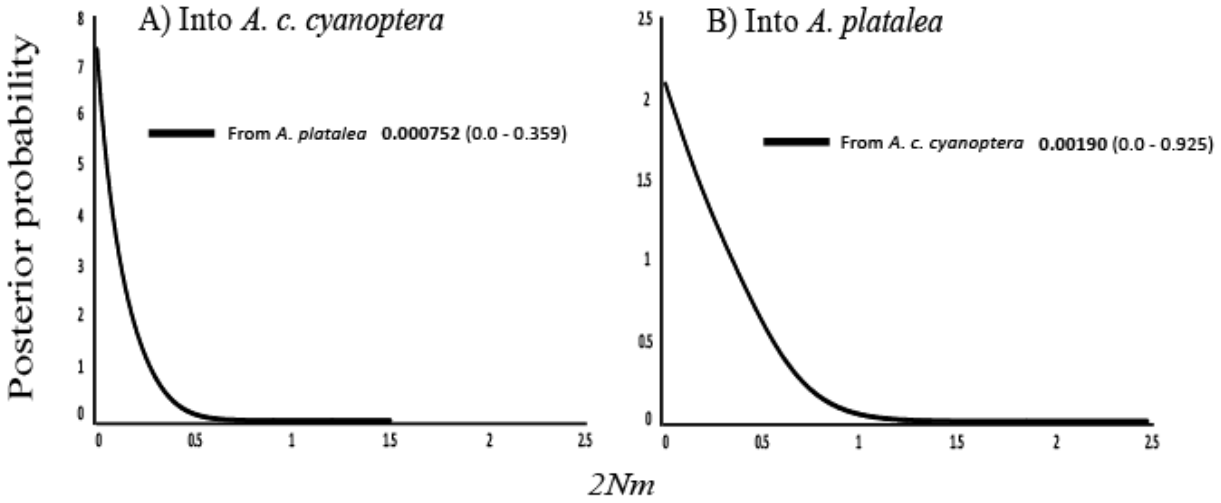

Figure 1-7. Estimates of the number of immigrants per generation for each South American comparison: *A. c. cyanoptera* (A) and *A. platalea* (B). Values in bold represent the highest posterior probability and the values in parentheses are the 95% highest posterior densities.

Table 1-1. Pairwise F_{ST} values among eight taxa of blue-winged ducks. Mitochondrial DNA above the diagonal, nuclear DNA (average among five loci) below the diagonal.

	A. rynchotis	*A. smithii*	*A. clypeata*	*A. platalea*	*A. c. cyanoptera*	*A. c. orinomus*	*A. c. septentrionalium*	*A. discors*
A. rynchotis	-	0.7922	0.5358	0.8741	0.8258	0.8489	0.8062	0.8328
A. smithii	0.4086	-	0.7872	0.9326	0.8993	0.9227	0.8866	0.9051
A. clypeata	0.0967	0.2434	-	0.8633	0.8113	0.8388	0.7895	0.8204
A. platalea	0.3213	0.2802	0.2946	-	0.8868	0.9131	0.8794	0.8965
A. cyanoptera								
cyanoptera	0.3086	0.4072	0.2665	0.3025	-	0.0723	0.4137	0.2527
orinomus	0.2572	0.3225	0.1815	0.2962	0.0972	-	0.4768	0.4175
septentrionalium	0.2023	0.2662	0.1343	0.2824	0.0796	0.0531	-	0.4823
A. discors	0.1864	0.2415	0.1019	0.3061	0.1435	0.0699	0.02334	-

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U. S. Government

Chapter 2. Low incidence of gene flow fails to counter genomic differentiation among allopatric species of a globally distributed lineage of blue-winged ducks

INTRODUCTION

Mechanisms of speciation vary among lineages, and the interactions among genetic drift, gene flow, and natural selection can have various contributions to patterns of genetic differentiation across the genome. Traditionally, allopatric speciation, or the lack of population connectivity, was thought to be the predominant mode of speciation (Mayr *et al.* 1963), and is expected when populations have disjunct distributions that are separated by dispersal barriers, such as mountain ranges, oceans, and deserts. It was thought that the origin of reproductive barriers must occur in allopatry for speciation to be complete (Dobzhansky & Dobzhansky 1937; Mayr 1942; Mayr *et al.* 1963; Mayr 2004). However, if diverging populations come back into contact (i.e., secondary contact) and are able to hybridize and produce viable offspring, then speciation was often considered incomplete due to the lack of reproductive isolation (Feder *et al.* 2012; Mayr 1942; Mayr *et al.* 1963; Dobzhansky & Dobzhansky 1937; Mayr 2004). The conditions for allopatric speciation are considered 'strict', where geographic barriers must be sufficient to prevent the homogenizing effects of gene flow. In speciation theory, if gene flow between diverging populations persists, then an allopatric speciation model no longer applies, and instead, models of parapatry or sympatry apply (Bird *et al.* 2012; Feder *et al.* 2012; Via 2012). These models of speciation are well theorized; however, they are more difficult to demonstrate in nature than allopatric speciation, and as a result, have been relatively neglected during the modern synthesis (Dobzhansky & Dobzhansky 1937; Mayr 1942; Mayr *et al.* 1963).

During the last few years, studies have begun to rigorously examine sympatric/parapatric models of speciation (i.e., divergence-with-gene-flow) and suggest that gene flow may be more important than previously recognized (Bierne *et al.* 2013; Bird *et al.* 2012; Cruickshank & Hahn 2014; Feder & Nosil 2010; Feder *et al.* 2012; Ferchaud & Hansen 2016; Flaxman *et al.* 2013; Harr 2006; Hohenlohe *et al.* 2010; Keller *et al.* 2013; Michel *et al.* 2010; Nadeau *et al.* 2012; Nosil 2008; Nosil & Feder 2012; Peters *et al.* 2012a; Saint-Laurent *et al.* 2003; Turner & Hahn 2010; Via 2012; White *et al.* 2010). Instead of relying on the absence of population connectivity, divergence-with-gene-flow models invoke an antagonistic relationship between divergent selection and gene flow (Cruickshank & Hahn 2014; Feder *et al.* 2012; Via 2012). Between populations, divergent selection can drive species divergence among a few loci while gene flow homogenizes regions that are not under selection or are not tightly linked to regions under selection (Feder *et al.* 2012; Via 2012). This interaction between divergent selection and gene flow creates a mosaic, or a 'semi porous' genome, and is most often studied between populations with overlapping distributions or very little geographic distance between sampling sites (Stölting *et al.* 2013). In contrast, most species with disjunct distributions are thought to be diverging under strict allopatry where geographic distance impedes gene flow between populations (Feder *et al.* 2012). However, this may not be the case for species with high dispersal tendencies. Specifically, species with the ability to disperse over large distances, such as those that undergo long-distance migrations, may facilitate gene flow among distant regions and result in patterns of divergence-with-gene-flow even if their distributions are seemingly allopatric (Proches & Ramdhani 2013).

Among highly dispersive lineages, the genus *Anas* (i.e., dabbling ducks) is particularly predisposed for global distributions (Lavretsky et al. 2014; Proches & Ramdhani 2013b; Peters *et*

al. 2012b), which is likely attributable to their vagile tendencies. Ancestral state reconstructions of geographic ranges suggest that dispersal-driven speciation is an important mechanism in this group (Johnson & Sorenson 1999). In brief, most species of dabbling ducks have as their closest relative a species from a different continent, suggesting that dispersal and colonization of new areas has contributed to speciation (Johnson & Sorenson 1999), and Proches and Ramdhani (2013) suggest that high vagility probably contributed to the global distributions of some lineages (i.e., cosmopolitan lineages). Although colonization of new continents would seem to set the stage for allopatric speciation, the high dispersal ability of ducks should also facilitate gene flow, even among geographically distant populations. Furthermore, waterfowl, and ducks in particular, are well known for their ability to hybridize and produce fertile offspring with both closely and more distantly related species (Tubaro & Lijtmaer 2002). High dispersal abilities and high propensities for hybridization may lead to complex patterns of genetic divergence in this group, where some species may be diverging in strict allopatry whereas others are diverging in the presence of gene flow.

In this study, I focused on a sub-lineage of dabbling ducks from the blue-winged duck clade, the shovelers. The shovelers comprise four species, the northern shoveler (*Anas clypeata*) distributed in North America, Europe, and Asia, the Australasian shoveler (*A. rhynchotis*) distributed in Australia and New Zealand, the Cape shoveler (*A. smithii*) distributed in southern Africa, and the red shoveler (*A. platalea*) distributed in South America. Given this global distribution, Proches and Ramdhani (2013) classified the shoveler group as one of 83 cosmopolitan avian lineages. Whereas the southern hemisphere species are mostly sedentary or undergo short distance movements, the northern shoveler undergoes seasonal long-distance migrations, which might facilitate gene flow among allopatric populations. Indeed, northern

shoveler vagrants have been observed within or near the range of both Australasian and Cape shovelers (see Chapter 1). In Chapter 1, a combination of nuclear DNA (nuDNA) intron and mitochondrial DNA (mtDNA) sequences revealed significant evidence of gene flow between at least some pairs of shovelers, whereas others seemed to be strictly allopatric. However, this study only examined six loci, which potentially provided low power for distinguishing between alternative models of divergence. Here I used Next-Generation-Sequencing (NGS) techniques to increase genomic coverage to provide more power and more resolution for examining evolutionary histories and for identifying the different patterns of speciation. In particular, I aimed to estimate levels of interspecific gene flow and to examine the frequency distributions of fixation indices to quantitatively differentiate between models of strict allopatry versus divergence with gene flow.

In the case of divergence-with-gene flow, an antagonistic relationship between divergent selection and gene flow exists and results in heterogeneous divergence throughout the genome (Feder *et al.* 2012). Under this model, I predict to detect levels of interspecific/intercontinental gene flow that are sufficient to offset the effects of genetic drift in driving substantial population divergence (>1 migrant per generation; (Slatkin 1987)). In addition, I expect to observe distributions of fixation indices (across all loci) that are highly skewed or L-shaped, containing many loci with low fixation indices and a few loci with high fixation indices (Feder *et al.* 2012). Such a distribution would indicate high heterogeneity in genomic differentiation between species pairs. Finally, I also expect to detect loci that do not fit neutral expectations (i.e., outlier loci, suggestive of a role of divergent selection). I argue that, if divergence-with-gene-flow is occurring in this group all three conditions must be met. Alternatively, under an allopatric model, gene flow is expected to be absent or insufficient to offset drift (<1 migrant per generation), the

distribution of fixation indices should be more uniform and not highly skewed, and few, if any, outlier loci should be detected (Nosil 2008; Cruickshank & Hahn 2014; Feder *et al.* 2012). To test these predictions within the shoveler complex, I used double-digest restriction-associated DNA sequencing (ddRAD-seq) to generate data for thousands of loci distributed throughout the genome.

METHODS

Sample collection:

A total of 88 samples were collected from five different continents: North America and Asia (northern shoveler, $N = 21$ and $N = 14$, respectively), South America (red shoveler, $N = 20$), Africa (Cape shoveler, $N = 17$), and Australia and New Zealand (Australasian shoveler, $N = 7$ and 9, respectively). Genomic DNA was extracted from tissue or blood using a DNeasy Blood & Tissue kit (Qiagen, Valencia, CA), and samples were quantified using a nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific Inc.) to ensure that all extracts had a minimum concentration of 10ng/µL.

RAD library preparation and assembly:

Sample preparation for generating ddRAD-seq libraries followed protocols outlined in DaCosta and Sorenson (2014). Specifically, 0.5 – 1.0 µg of genomic DNA was digested using two restriction enzymes, SbfI and EcoRI. Barcoded adaptors were then ligated to the sticky ends of each DNA fragment; each individual received a unique barcode-adapter combination. Adapter-ligated DNA were then size-selected using gel electrophoresis (2% low-melt agarose) and were extracted using a MinElute gel extraction kit (Qiagen). I selected fragments of 300-450 bp (including adapters), although smaller fragments (~150 bp) are also captured using this

method (see (DaCosta & Sorenson 2014). Size-selected DNA fragments were then amplified using PCR and Phusion high-fidelity DNA polymerase (Thermo Scientific, Pittsburgh, PA, USA). PCR products were purified using AMPure magnetic beads (Beckman Coulter, Inc., Indianapolis, IN). Cleaned products underwent a final quality check using real-time PCR and pooled in equimolar concentrations. Pooled libraries were sequenced on an Illumina HiSeq2500 Sequencer at Tufts University.

Sequence data were de-multiplexed and assembled using the pipeline constructed by DaCosta and Sorenson (2014), which includes custom Python scripts available at http://github.com/BU-RAD-seq/ddRAD-seq-Pipeline. For each individual, low-quality reads (any read that had a quality score < 30) were filtered, and identical reads were collapsed while maintaining read count and the highest quality score at each position. Reads were then clustered into putative loci using the UCLUST port in USEARCH v.5 (Edgar 2010) and an –id setting of 0.85 (85% pairwise identity). I mapped the highest quality read from each cluster to the Mallard reference genome (accession numbers SS263068950 – SS263191362; (Kraus *et al.* 2011; Huang *et al.* 2013) using BLASTN v. 2 (Altschul *et al.* 1990), and clusters with identical BLAST hits ($\pm$ 50 bp on the same reference scaffold) were combined. Reads within each cluster were aligned using MUSCLE v. 3 (Edgar 2004), and each individual was genotyped at each cluster using the Python script *RADGenotypes.py*. Overall, homozygous genotypes were scored if >93% of sequence reads were consistent with a single haplotype, whereas heterozygotes were scored if a second haplotype was represented by $\geq$29% of reads (see DaCosta & Sorenson 2014). In addition, individuals were scored as heterozygous if a second allele was represented by 20–29% of reads and was found in other sampled individuals. Clusters with end gaps (sequence overhang) due to indels and/or a polymorphism in the SbfI restriction site were either

automatically trimmed or flagged for manual editing. In addition, clusters that contained ≥ 2 polymorphisms in the last five base pairs or >19 polymorphic sites overall were flagged. Manual checking and editing of flagged loci was done in Geneious (Biomaters Inc., San Francisco, CA, USA). I retained all loci that contained <20% missing genotypes and contained ≤5% flagged genotypes after manually editing clusters.

I partitioned all recovered clusters into sex-linked on the Z-chromosome and autosomal loci. All captured sequences were mapped to the mallard genome that has been assembled into chromosomes (T. Faraut, unpubl. data). Loci that had a single match with >85% identity to an autosome or the Z chromosome were categorized accordingly. However, there were few loci that either did not have a BLAST hit or had multiple hits to the mallard genome. For these loci I used overall male and female heterozygosity and sequence depth ratios to identify Z-linked and autosomal loci (Lavretsky et al. 2015). Because female birds are the heterogametic sex (ZW), Z-linked markers will have no heterozygosity (i.e., they will be hemizygous) and half the depth compared to autosomal loci. Based on the observed ratio of male depth/female depth of loci mapped to known chromosomes, I used a depth ratio of < 0.97 to categorize unmapped loci as autosomal and > 1.27 (with <10% of females being heterozygous to allow for minor sequence errors) for Z-loci. If the depth ratio of unmapped loci fell between 0.97 and 1.27, I counted all loci as autosomal if >5% of females were heterozygous; otherwise, the loci were categorized us "unknown".

Estimates of gene flow and evolutionary histories:

I used the Bayesian clustering method in the program SplitsTree v4.13 which provides a framework for evolutionary analysis by combining both phylogenetic trees and networks to visualize genetic variation within and among populations (Ayling & Brown 2008). Combining

autosomal and Z-linked loci, I constructed a splits tree using consensus sequences of my concatenated nuDNA (with heterozygous sites coded with IUPAC ambiguity codes). I used P-distances with ambiguities treated as the average state (e.g., T vs. C is a distance of 1.0, whereas Y vs. C is a distance of 0.5) to determine the proportion of differences per site.

To estimate gene flow among the shoveler species, I used the Generalized Phylogenetic Coalescent Sampler G-PhoCS v.1.2.3 (Gronau *et al.* 2011). Specifically, G-PhoCS is a coalescent program that is used for inferring ancestral and contemporary population sizes, population divergence times, and migration rates from genomic data (Gronau *et al.* 2011). I defined all priors as the default setting, which included a gamma distribution for each parameter. I ran a Markov Chain with a burn-in of 200,000 for 1,000,000 iterations sampling parameters every 10 iterations with a constant per locus mutation rate resulting in a total of 100,001 samples. My analysis included a total of 22 parameters, including seven parameters of population size (four for contemporary populations and three for ancestral populations), three parameters for divergence times, and 12 migration parameters (accounting for bidirectional migration across all species pairs). Following the phylogenetic tree from chapter one, I defined northern and Australasian shovelers as sister taxa, the Cape shoveler as sister to the northern-Australasian ancestor, and the red shoveler as sister to the northern-Australasian-Cape ancestor. Because G-PhoCS is computationally intensive, following Campagna et al. (2015), I subsampled individuals by randomly selecting four individuals per species for coalescent analyses. Large genomic data sets permit sampling fewer individuals without losing confidence in parameter estimates (Campagna *et al.* 2015). Additionally, to calculate the per site mutation rate for my coalescent analyses, I mapped all empirical loci to the swan goose genome (*Anser cygnoides*) and calculated the number of pairwise differences per site (2,992 loci were mapped to the goose genome). I

assumed a divergence date between ducks and geese of 20.8 million years (Lu *et al.* 2015), and used $d = 2\mu T$ (where d is the number of pairwise difference per site, μ is the per site mutation rate, and T is time since divergence measured in years) to calculate the per site mutation rate for ddRAD-seq loci.

Genomic differentiation:

To assess genomic differentiation among the four shoveler species, I used PopGenome (Pfeifer *et al.* 2014), which is an R-package (R Development Core Team 2013) for genomic analysis. Specifically, I estimated genomic differentiation as Φ_{ST} (the proportion of the total nucleotide diversity partitioned between populations) for each locus for each pairwise comparison (six comparisons total). I then constructed Φ_{ST} frequency distributions by binning Φ_{ST} values in 0.01 increments.

Detecting outlier loci and neutral simulations:

To identify any deviations from neutral expectations (outlier loci), I used the Bayesian simulation method of Beaumont & Balding (2004) as implemented in BAYESCAN v.2.1 (Foll & Gaggiotti 2008). Because previous studies have shown that simulations that account for a deep coalescence (increased prior odds) have a lower probability of producing false-positives within a larger data set (Lotterhos & Whitlock 2014; Lotterhos & Whitlock 2015), I ran each analysis with prior odds set to 100, and conducted 20 pilot runs with a run length of 5,000 iterations each, followed by a burn-in of 50,000 steps and 5,000 sampling steps. I ran BAYESCAN for each of the six pairwise comparisons. I identified all significant outlier loci with a false discovery rate (FDR, the proportion of outliers that are deemed as false-positives) of 0.01 (p-value = 0.01).

As a second test for outlier loci that incorporates evolutionary histories, I simulated genetic diversity following the methods described by Peters et al. (2012). To simulate Φ_{ST} distributions under neutral evolution I used the program MS (Hudson 2002) and generated multiple independent loci under a four-population isolation-with-migration model. Parameters for MS were taken from the coalescent analysis (G-PhoCS), including t (where $t = T/\mu$, and T is the number of years since divergence), Θ (where $\Theta = 4N_e\mu$, and N_e is the effective population size and μ is the mutation rate per site per generation), and $2Nm$ (where, N is the population size, and m is the proportion of the population consisting of immigrants per generation). Following Peters et al. (2012), Θ for northern shovelers was converted to the per-locus mutation rate, and all other parameters were rescaled to the $4N_e$ of northern shovelers. To mimic empirical data, I generated a total of 3,248 independent loci consisting of 88 individuals each ($2N$ = 176 simulated sequences per locus). For each simulated locus, I adjusted parameter values for varying locus lengths in the empirical data and for differences in modes of inheritance between autosomal and Z-linked loci (see Peters et al. 2012 for detailed explanations of these adjustments).

RESUTS

Genomic assembly and recovered loci

Overall, I obtained an average of 760,991 (range = 305,999 – 4,994,016) high quality sequence reads per individual. After filtering the data, I retained 3,248 loci that were recovered from >80% of individuals and contained $\leq$5% flagged genotypes. Among these loci, the median depth was 119 reads per individual per locus, and genotypes were complete for 98% and partial (one allele was scored) for 0.9% of individuals per locus. The final data set included 3,152 loci with at least one polymorphism and 96 constant loci, and a total of 404,413 aligned nucleotides

containing 25,299 SNPs. Based on BLAST hits to the assembled chromosomes of the mallard and patterns of male/female read depth and heterozygosity, I inferred that 3,071 loci (2,985 variable loci) were autosomal and 177 (167 variable loci) were Z-linked. Overall nucleotide diversity was highest for the northern shoveler ($\pi = 0.0069$), intermediate for the red shoveler ($\pi = 0.0056$) and Australasian shoveler ($\pi = 0.0056$), and lowest for the Cape shoveler ($\pi = 0.0019$).

Neighbor-net trees generated from a concatenated consensus sequences showed clear clustering of each shoveler species (Fig. 2-1). Specifically, each individual clustered more closely with other individuals from the same species rather than individuals of the other species. The Cape shoveler showed the tightest clustering consistent with low within-population diversity, whereas the northern shoveler cluster was more dispersed and consistent with the higher observed level of within-population diversity. Within the Australasian shoveler cluster, I recovered two different sub-groups, one consisting of samples from Australia and the other consisting of samples from New Zealand. In contrast, the two populations of the northern shoveler (Eurasia vs. North America) were broadly intermixed. Furthermore, consistent with species-tree estimates from mtDNA and five introns (chapter 1), the northern and Australasian samples clustered more closely with each other than with the other species.

Evolutionary histories inferred from coalescent analysis:

Results from coalescent analysis, implemented in G-PhoCS, suggested a large range of effective population sizes, divergence times, and migration rates among species. Specifically, the northern shoveler had the largest effective population size ($\Theta = 0.013$), followed by the red shoveler ($\Theta = 0.0047$), the Australasian shoveler ($\Theta = 0.0019$), and the Cape shoveler ($\Theta = 0.00042$) (Fig. 2-2a). Additionally, estimates of Θ for ancestral populations suggested that the most recent common ancestor between the northern and Australasian shoveler ($\Theta = 0.0037$) had

a smaller population size than the northern shoveler whereas confidence intervals overlapped the Θ for the Australasian shoveler (Fig. 2-2a). In contrast, estimates of Θ for the most recent common ancestor between northern, Australasian, and Cape shoveler suggested a population size (Θ = 0.015) equal to or larger than the contemporary northern shovelers. Estimates of Θ for the 'ancestral shoveler' (most recent common ancestor of all shoveler species) suggested a larger population size (Θ = 0.0071) compared to contemporary southern hemispheric shovelers but smaller than the northern shoveler. Thus, estimates of ancestral effective population sizes suggest large fluctuations over time.

Estimates of divergence between the northern and Australasian shoveler ($\tau = 0.00042$, T = 420,000 years before present, ybp) were only slightly more recent than between the northern/Australasian and Cape shoveler ($\tau = 0.00044$, T = 440,000 ybp), and confidence intervals were broadly overlapping. In contrast, estimates from the initial split within the 'ancestral shoveler' suggested a much deeper divergence time ($\tau = 0.00175$, T = 1,754,000 ybp) (Fig. 2-2b).

Estimates of migration rates suggested less than one migrant per generation ($2Nm$) between all pairs of shovelers. However, six of the 12 estimates had confidence intervals that included the lowest bin suggesting that the data were consistent with no gene flow. Furthermore, three of the estimates were very close to $2Nm = 0$, suggesting that gene flow has been extremely rare ($2Nm < 0.1$ migrants per generation) (Fig. 2-2c). However, migration rates were consistently higher from the northern shoveler into the southern hemisphere shoveler species than in the reverse direction. Estimates of migration from northern into Australasian shoveler yielded 0.88 migrants per generation, northern into red shoveler yielded 0.65 migrants per generation, and northern into Cape shoveler yielded 0.036 migrants per generation (Fig. 2-2c). In contrast, gene

flow in the reverse direction was estimated to be 0.08, 0.37 and 0.00091, respectively. These results also suggested non-zero levels of bi-directional gene flow between Cape and red shoveler, although estimates were less than 0.1 migrant per generation in both directions.

Genomic differentiation:

Φ_{ST} values were similar between autosomal and Z-linked loci, although Φ_{ST} was slightly higher for Z-linked loci compared to autosomal loci (Table 2-1) which is expected given that the Z-chromosome has three-fourths the effective population size of autosomal DNA. Combining marker types, Φ_{ST} ranged between 0.10 and 0.53; the northern and Australasian shoveler were the least differentiated (Φ_{ST} = 0.10), whereas the Cape and red shoveler were the most differentiated (Φ_{ST} = 0.53) (Table 2-1). In general, pairwise comparisons with the Cape shoveler yielded the highest measures of genomic differentiation and pairwise comparisons with the northern shoveler yielded the lowest measures of genomic differentiation.

For all empirical Φ_{ST} distributions, skewness values ranged from 0.60 to 2.27. Specifically, all comparisons against the Cape shoveler had the most symmetrical distributions ranging from 0.41 to 0.94, and all comparisons against the northern shoveler yielded the highest skewness values ranging from 0.94 to 2.19 (Fig. 2-3). The higher skewness in comparisons with the northern shoveler, especially between the northern and Australasian shovelers (skewness = 2.19), reflected the strong peak near Φ_{ST} = 0 and a prominent tail extending toward higher Φ_{ST}'s creating an "L-shaped" distribution.

Detecting outlier loci and neutral simulations:

Φ_{ST} distributions that were simulated under models of neutral evolutionary history were qualitatively similar to empirical distributions for all pairwise comparisons (Fig. 2-3). Overall

Φ_{ST} ranged from 0.14 – 0.53 and were similar to empirical Φ_{ST} values. Specifically, comparisons involving the northern shoveler yielded the lowest levels of genomic differentiation and comparisons against the Cape shoveler yielded the highest Φ_{ST} values (Fig. 2-3). Although simulated Φ_{ST} distributions were slightly less skewed (0.378 – 1.60) than empirical distributions (0.41 – 2.19), pairwise comparisons that included the Cape shoveler yielded the most symmetrical distributions in both data sets, whereas comparisons with the northern shoveler yielded the most asymmetrical (positively skewed) distributions (Fig. 2-3). In two of the six pairwise comparisons (northern vs. red shoveler and Australasian vs. red shoveler), empirical Φ_{ST} distributions exceeded the range of neutral simulations suggesting the possible role of divergent selection for 85 loci in total (Fig. 2-3). In the comparison between the red shoveler and northern shoveler and between the red and Australasian shoveler, 40 loci and 45 loci fell outside the simulated values, respectively. Thirty-two loci exceeded neutral expectations in both comparisons. In all other comparisons, the empirical values of Φ_{ST} fell within the simulated range for all loci.

Analyses in BAYESCAN 2.1 identified a total of 36 outlier loci. Thirty-three loci were located on the autosomes and three were located on the Z-chromosome (Fig. 2-4ade). Comparisons between the red and northern shoveler yielded 30 outliers, and comparisons between the red and Australasian shoveler yielded five outliers; three loci were outliers in both comparisons. In addition, the northern and Australasian shoveler comparison yielded only one outlier locus, although there were two additional loci that fell just outside of my thresholds for calling outlier loci (P = 0.011 and 0.013, respectively).

DISCUSSION

Sequences from more than 3,000 loci have provided limited evidence that the globally distributed shoveler lineage is diversifying under a model of divergence-with-gene-flow. First, although six of the twelve migration parameters had confidence intervals that did not overlap zero, suggesting that there has been some gene flow during speciation, all estimates suggested less than one migrant per generation. This magnitude of gene flow is likely insufficient to counter the effects of genetic drift (Slatkin 1987), permitting genetic divergence throughout the genome with or without selection. Indeed, I found high levels of genomic differentiation among all four species within the shoveler lineage. Second, skewness values for the empirical Φ_{ST} distribution was only slightly wider than the range for the distribution simulated under a neutral evolutionary history. Third, although neutral simulations and F_{ST} outlier tests detected a number of outlier loci, 25 of which were identified using both methods, the occurrence of outlier loci does not appear to be associated with species pairs that have low background levels of differentiation and high estimates of gene flow. For example, under a divergence-with-gene-flow model, outlier loci are expected to be most prominent between the northern and Australasian shovelers (lowest Φ_{ST}/highest gene flow). However, most statistical outliers were found between the red shoveler and the northern/Australian shovelers (high Φ_{ST}/low-intermediate levels of gene flow). Overall, these data support limited gene flow that is insufficient for offsetting drift and that models of allopatric speciation, albeit not strictly, best describe shoveler diversification.

Gene flow and evolutionary history

I detected six species comparisons where 95% confidence intervals did not overlap zero, suggesting gene flow. Importantly, most of the gene flow seems to have originated from the northern shoveler, as I found non-zero estimates of gene flow from the northern shoveler into every other shoveler species. In comparison to mtDNA and five nuclear introns (chapter 1),

ddRAD-seq seems to have provided more power for rejecting strict allopatry. Although the combination of mtDNA and nuclear introns supported unidirectional intercontinental gene flow from northern into Australasian shoveler and from Cape into Australasian shoveler, there was insufficient resolution to reject scenarios of no gene flow. In contrast, analysis of ddRAD-seq loci supported non-zero values of gene flow for six species comparisons. Furthermore, whereas mtDNA and introns suggested gene flow from the Cape shoveler into the Australasian shoveler (albeit with wide confidence intervals), ddRAD-seq strongly supported a scenario of no gene flow between these species.

Finding non-zero estimates of gene flow from the northern shoveler into all other shoveler species was not surprising given the large effective population size and migratory behavior of the northern shoveler. The highest estimates of gene flow were between the northern and Australasian shoveler (~ 0.88 migrants per generation), and this pattern was also observed in chapter 1 based on analysis of mtDNA and five nuclear introns. Furthermore, Australasian shovelers in New Zealand have mtDNA haplotypes that seem to be derived from the northern shoveler, providing strong support for gene flow (Chapter 1). The northern shoveler has expanded its migration range into the Philippines and Malaysia (Van Weerd & Van der Ploeg 2004), and vagrant individuals have been observed in Australia and New Zealand (Close & Jaensch 1981; Marchant & Higgins 1990), and reported in public databases, such as eBird (ebird.org;(Sullivan *et al.* 2009) and the Atlas of Australian Birds (birddata.com.au).

Surprisingly, however, I also found non-zero estimates of gene flow among sedentary species, particularly between the Cape and red shovelers. Because the Cape shoveler is endemic to the southern portion of Africa and experiences no seasonal migration, and the red shoveler is endemic to southern South America and undergoes only short-distance migration, gene flow

seems unlikely. Regardless, estimates of gene flow in both comparisons suggested fewer than 0.1 migrants per generation, and therefore, might reflect the occurrence of rare vagrants (i.e., sweepstake events; (Simpson 1940)) moving across the Atlantic Ocean. Several non-migratory waterfowl species are distributed in both South America and southern Africa (e.g., comb duck, *Sarkidiornis melanotos*; southern pochard, *Netta erythrophthalma*; fulvous whistling duck, *Dendrocygna bicolor*; white-faced whistling duck, *D. viduata*), illustrating the importance of sweepstakes events in contributing to biodiversity in the southern hemisphere. Alternatively, it is possible that migration estimates reflect bouts of ancestral gene flow. Because GPhocS is computationally intensive, it was not possible to include ancestral migration parameters in my analyses, and therefore, any signatures of ancestral gene flow would have to be incorporated into estimates of contemporary gene flow.

Genomic distribution

Under models of divergence-with-gene-flow, I would expect a positively skewed "L-shape" distribution with a small proportion of the genome being strongly differentiated, whereas the majority of loci would be weakly differentiated (Feder *et al.* 2012; Lemay & Russello 2015). In contrast, for models of allopatry, I would expect more symmetrical distributions with a larger proportion of highly differentiated loci resulting from the stochastic effects of genetic drift (Feder *et al.* 2012). Out of the six pairwise comparisons, the northern/Australasian shoveler, red/northern shoveler, and the red/Australasia shoveler comparisons had the largest skewness values yielding "L-shaped" distributions with few highly differentiated loci. This pattern is consistent with other studies where the homogenizing effects of gene flow and the divergent effects of selection cause distributions to be highly asymmetrical (Feder & Nosil 2010; Lemay & Russello 2015; Ruegg *et al.* 2014). Given my coalescent parameters, neutral simulations suggest

that demography may play a larger role than the interactions between gene flow and selection in the size and shape of Φ_{ST} distributions. Specifically, my results indicate that neutral demographic history (i.e., genetic drift) can cause patterns in Φ_{ST} distributions that appear to be consistent with models of divergence-with-gene-flow. For example, under models of divergence-with-gene-flow, I expect Φ_{ST} distributions to be positively skewed with a prominent 'tail' due to the interaction between divergent selection and gene flow. However, here I show that regardless of the interaction between divergent selection and gene flow, demographic histories such as effective population size and time since divergence can generate distributions that mimic those expected under divergence-with-gene-flow. For instance, on average, comparisons against the northern shoveler yielded the highest skewness values with an extended tail (high skewness) suggesting a large portion of weakly to moderately differentiated loci between species with few more strongly differentiated loci. Relatively high skewness was observed in both the empirical and the simulated data sets. In contrast, for species with smaller effective population sizes and larger times since divergence, skewness values were lower and Φ_{ST} distributions were flatter, consisting of mainly weakly, moderately, and strongly differentiated loci. This effect of demography can especially be seen in the distributions involving comparisons with the Cape shoveler, which had the smallest effective population size. Thus, the ratio between effective population size and time since divergence (N_e/T) is important to consider when identifying genomic signatures of speciation models.

Natural selection, outlier loci, and demography

Consistent with divergence with gene flow, I found evidence that selection contributed to divergence within the shovelers. Specifically, using BAYESCAN I recovered a total of 36 outlier loci, and results from neutral simulations suggested 85 outlier loci. Importantly, I found 25 loci

that were consistent between the two methods. However, outlier loci were only detected among three pairwise comparisons including northern/Australasian shoveler, northern/red shoveler, and Australasian/red shoveler. Two of these comparisons (northern/Australasian shoveler and northern/red shoveler) also had non-zero estimates of gene flow suggestive of a model of divergence-with-gene-flow where gene flow suppresses neutral divergence causing outlier loci to be more prominent (Feder *et al.* 2012; Via 2012; Via & West 2008). This scenario seems more probable between the northern and Australasian shoveler comparison because they are the least differentiated and experience the highest levels of gene flow. However, only one locus in the northern/Australasian shoveler comparison was recovered as an outlier, whereas most of the outliers were observed for the northern/red shoveler and Australasian/red shoveler.

The northern/red shoveler and the Australasian/red shoveler comparisons yielded estimates of low (or no) gene flow and high levels of genomic differentiation, yet outlier loci were prominent in both comparisons. It is possible that these outliers are linked to loci that are involved in local adaptation (i.e., ecological speciation; Nosil 2012), so that selection is driving changes in allelic frequencies regardless of the presence or absence of gene flow. Under this scenario, even weak selection over the long-term could contribute to divergent selection and the presence of detectable outlier loci. Alternatively, true demographic histories are likely much more complex (potentially including population expansions, population bottlenecks, and gene flow with other species) than the models analyzed in G-PhoCS, and this complexity could have contributed to the skewness of the Φ_{ST} distributions. Therefore, the presence of outlier loci might reflect more complex neutral demographic histories than what I simulated. Functional tests are needed to determine if selection has caused the observed "outlier loci". .

CONCLUSION

Traditionally, allopatric speciation was thought to be the predominant mode of speciation (Dobzhansky & Dobzhansky 1937; Mayr 1942; Mayr *et al.* 1963). In this case, geographic barriers such as deserts, distance, oceans, and mountain ranges restrict the movement of individuals and prevent the homogenizing effects of gene flow. Here, I show that in the case of a highly dispersive cosmopolitan lineage, geographic barriers that create disjunct distributions are not strong enough to cause strict allopatry among geographically separated species. However, the low migration rates are probably insufficient to offset divergence due to genetic drift. Therefore, patterns of speciation within the shovelers better fits a model of allopatry, despite the occurrence of gene flow and outlier loci. I recommend that future studies use caution when using Φ_{ST} distributions to infer models of speciation as I show that effective population size and time since divergence can influence the size and shape of these distributions. Finally, this study illustrates how allopatric speciation can serve as a null model when testing modes of speciation.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403-410.

Ayling SC, Brown TA (2008) Novel methodology for construction and pruning of quasi-median networks. *BMC bioinformatics*, **9**, 115-2105-9-115.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular ecology*, **13**, 969-980.

Bierne N, Gagnaire P, David P (2013) The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Cur. Zool.*, **59**, 72-86.

Bird CE, Fernandez-Silva I, Skillings DJ, Toonen RJ (2012) Sympatric speciation in the post "modern synthesis" era of evolutionary biology. *Evolutionary Biology*, **39**, 158-180.

Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ (2015) Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular ecology*, **24**, 4238-4251.

Close D, Jaensch R (1981) Northern Shoveler at Coongie Lake. *S.Aust.Ornith*, **28**, 178-179.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133-57.

DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol.

Dobzhansky T, Dobzhansky TG (1937) *Genetics and the Origin of Species* Columbia University Press.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England),* **26**, 2460-2461.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research,* **32**, 1792-1797.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics,* **28**, 342-350.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution,* **64**, 1729-1747.

Ferchaud A, Hansen MM (2016) The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: three- spine sticklebacks in divergent environments. *Molecular ecology,* **25**, 238-259.

Flaxman SM, Feder JL, Nosil P (2013) Genetic Hitchhiking and the Dynamic Buildup of Genomic Divergence during Speciation with Gene Flow. *Evolution,* **67**, 2577-2591.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics,* **180**, 977-993.

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nature genetics,* **43**, 1031-1034.

Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome research*, **16**, 730-737.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *Plos Genetics*, **6**, e1000862.

Huang Y, Li Y, Burt DW, et al (2013) The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nature genetics*, **45**, 776-783.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics (Oxford, England)*, **18**, 337-338.

Johnson KP, Sorenson MD (1999) Phylogeny and biogeography of dabbling ducks (genus: *Anas*): a comparison of molecular and morphological evidence. *Auk*, **116**, 792-805.

Keller I, Wagner CE, Greuter L, et al (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular ecology*, **22**, 2848-2863.

Kraus RH, Kerstens HH, Hooft P, et al (2011) Genome wide SNP discovery, analysis and evaluation in mallard (Anas platyrhynchos). *BMC genomics*, **12**, 1.

Lemay MA, Russello MA (2015) Genetic evidence for ecological divergence in kokanee salmon. *Molecular Ecology*, **24**, 798-811.

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular ecology*, **24**, 1031-1046.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular ecology*, **23**, 2178-2192.

Lu L, Chen Y, Wang Z, et al (2015) The goose genome sequence leads to insights into the evolution of waterfowl and susceptibility to fatty liver. *Genome biology*, **16**, 89-015-0652-y.

Marchant S, Higgins PJ (1990) Handbook of Australian, New Zealand and Antarctic birds; volume 1.

Mayr E (1942) *Systematics and the origin of species, from the viewpoint of a zoologist* Harvard University Press.

Mayr E, Mayr E, Mayr E, Mayr E (1963) *Animal species and evolution* Belknap Press of Harvard University Press Cambridge, Massachusetts.

Mayr E (2004) Happy birthday: 80 years of watching the evolutionary scenery. *Science (New York, N.Y.)*, **305**, 46-47.

Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9724-9729.

Nadeau NJ, Whibley A, Jones RT, et al (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B-Biological Sciences,* **367**, 343-353.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology,* **17**, 2103-2106.

Nosil, Patrik. *Ecological speciation*. OUP Oxford, 2012.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B-Biological Sciences,* **367**, 332-342.

Pérez-Figueroa A, García-Pereira M, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267-2276.

Peters JL, McCRACKEN KG, Pruett CL, et al (2012a) A parapatric propensity for breeding precludes the completion of speciation in common teal (Anas crecca, sensu lato). *Molecular ecology*, **21**, 4563-4577.

Peters JL, Bolender KA, Pearce JM (2012b) Behavioural vs. molecular sources of conflict between nuclear and mitochondrial DNA: the role of male-biased dispersal in a Holarctic sea duck. *Molecular Ecology*, **21**, 3562-3575.

Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ (2014) PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Molecular biology and evolution,* **31**, 1929-1936.

Proches S, Ramdhani S (2013a) Eighty-three lineages that took over the world: a first review of terrestrial cosmopolitan tetrapods. *Journal of Biogeography,* **40**, 1819-1831.

Proches S, Ramdhani S (2013b) Eighty-three lineages that took over the world: a first review of terrestrial cosmopolitan tetrapods. *Journal of Biogeography,* **40**, 1819-1831.

Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular ecology,* **23**, 4757-4769.

Saint-Laurent R, Legault M, Bernatchez L (2003) Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (Osmerus mordax Mitchill). *Molecular ecology,* **12**, 315-330.

Simpson GG (1940) Mammals and land bridges.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science (New York, N.Y.),* **236**, 787-792.

Stölting KN, Nipper R, Lindtke D, et al (2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular ecology,* **22**, 842-855.

Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling S (2009) eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, **142**, 2282-2292.

Tubaro PL, Lijtmaer DA (2002) Hybridization patterns and the evolution of reproductive isolation in ducks. *Biological Journal of the Linnean Society,* **77**, 193-200.

Turner TL, Hahn MW (2010) Genomic islands of speciation or genomic islands and speciation? *Molecular ecology,* **19**, 848-850.

Van Weerd M, Van der Ploeg J (2004) Surveys of wetlands and waterbirds in Cagayan valley, Luzon, Philippines. *Forktail,* **20**, 33-39.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular ecology,* **17**, 4334-4345.

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B-Biological Sciences,* **367**, 451-460.

White BJ, Cheng C, Simard F, Costantini C, Besansky NJ (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of Anopheles gambiae. *Molecular ecology,* **19**, 925-939.

Table 2-1: Genomic differentiation (ΦST) for all six pairwise comparisons for (a) autosomal loci (above the diagonal) and Z-linked loci (below the diagonal) and (b) ΦST for autosomal and Z-linked loci combined (empirical data; above the diagonal) and ΦST under neutral expectations (simulated data; below the diagonal).

(a)

	Australasian	Cape	Northern	Red
Australasian	-	0.49	0.10	0.32
Cape	0.53	-	0.40	0.53
Northern	0.15	0.48	-	0.24
Red	0.43	0.69	0.33	-

(b)

	Australasian	Cape	Northern	Red
Australasian	-	0.49	0.10	0.33
Cape	0.52	-	0.40	0.54
Northern	0.14	0.42	-	0.25
Red	0.30	0.54	0.21	-

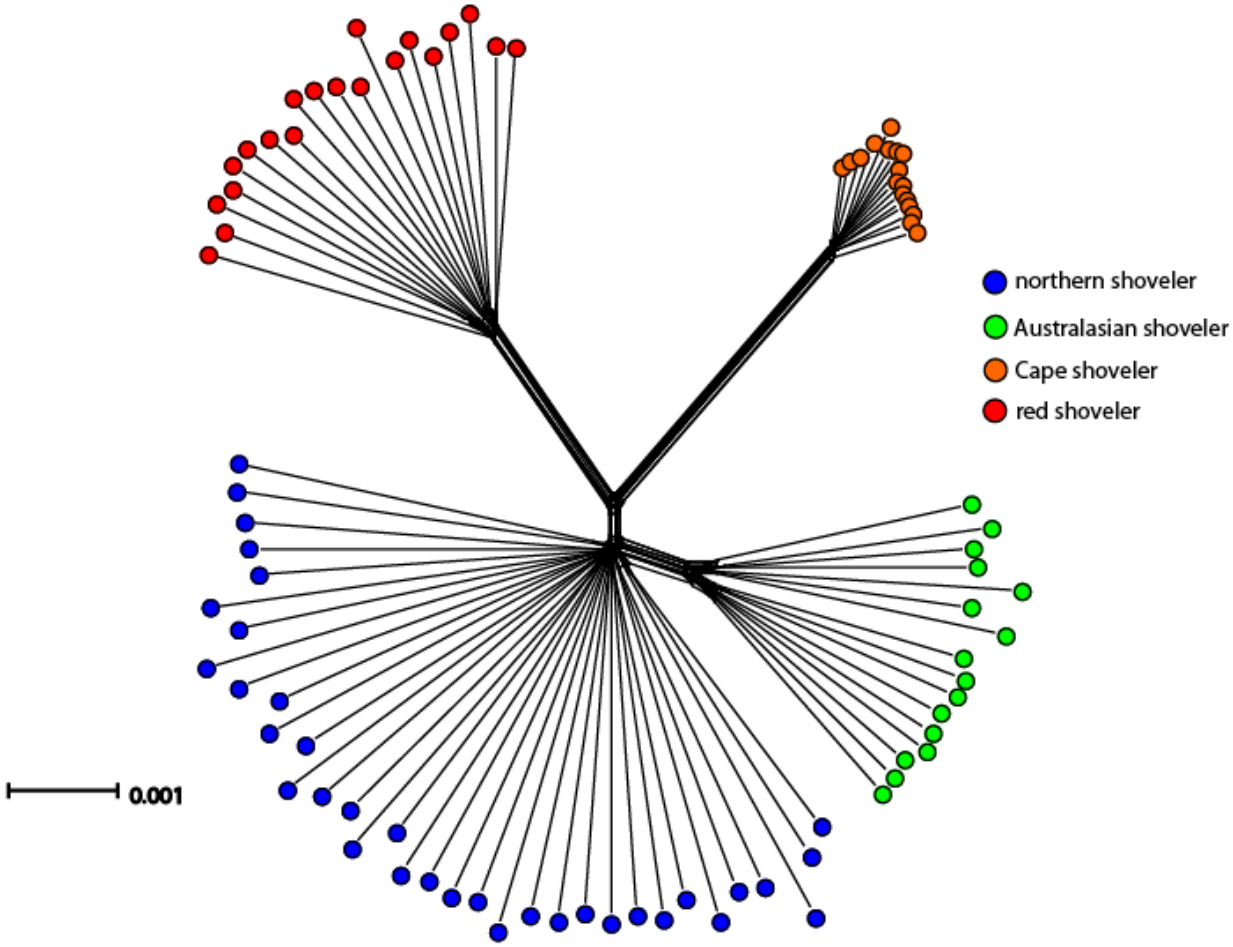

Figure 2-1. Neighbor-net tree of the shovelers constructed from 3,248 concatenated loci. Each species is denoted by a different color: northern shoveler, blue; Australasian shoveler, green; Cape shoveler, orange; and red shoveler, red. Note that the branch lengths suggests that the Cape shoveler is the most differentiated species and also has the least genetic diversity. In contrast, the northern shoveler has the greatest genetic diversity.

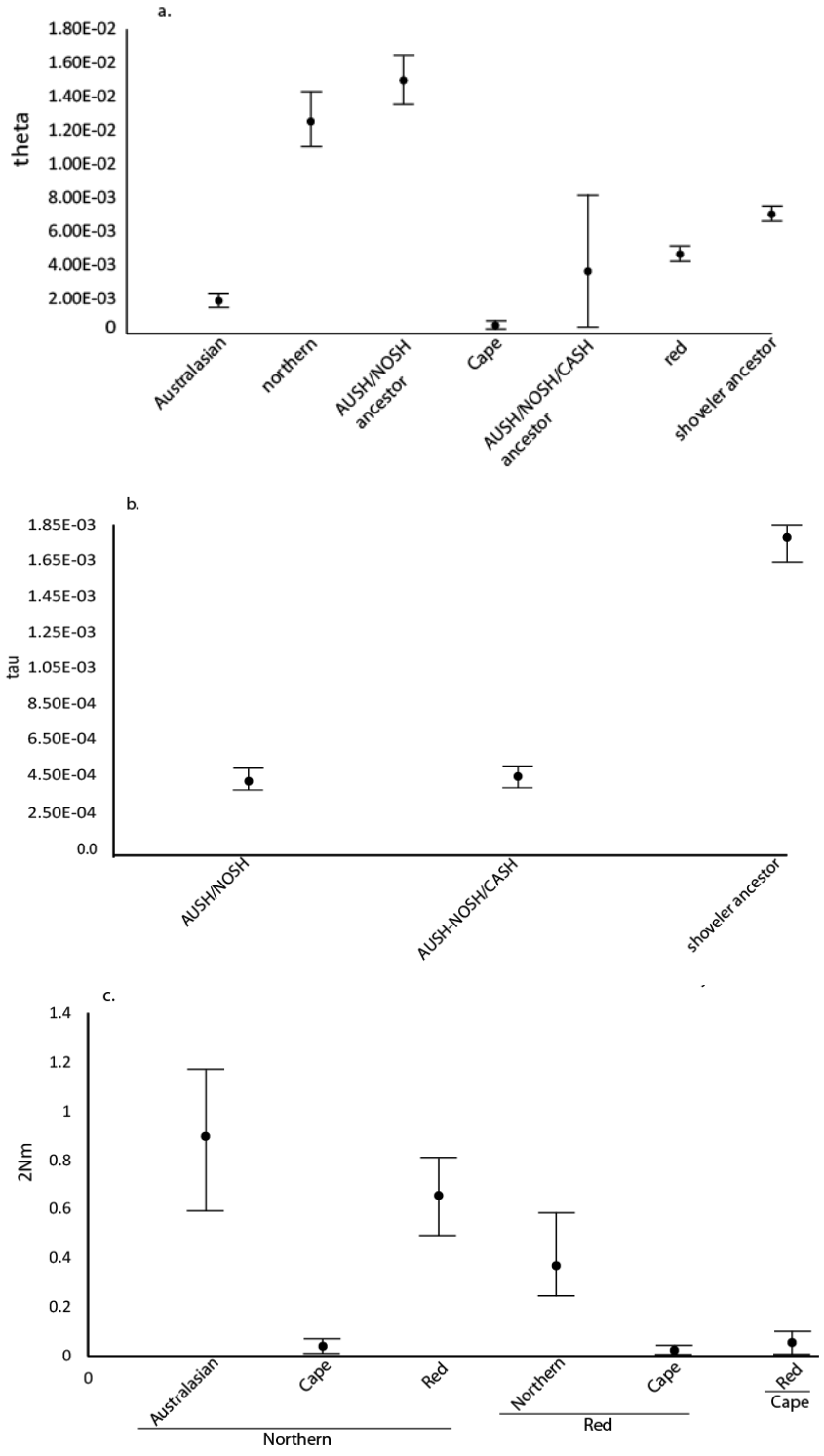

Figure. 2-2: Results from my coalescent analysis implemented by G-PhoCS, where a) is the estimated effective population size (in the form of theta) for all contemporary and ancestral species, b) the time of each speciation event within the shovelers, and c) the six nonzero estimates of gene flow (in the form of 2Nm). The 95% confidence intervals are denoted by the horizontal bar at the end of each vertical bar.

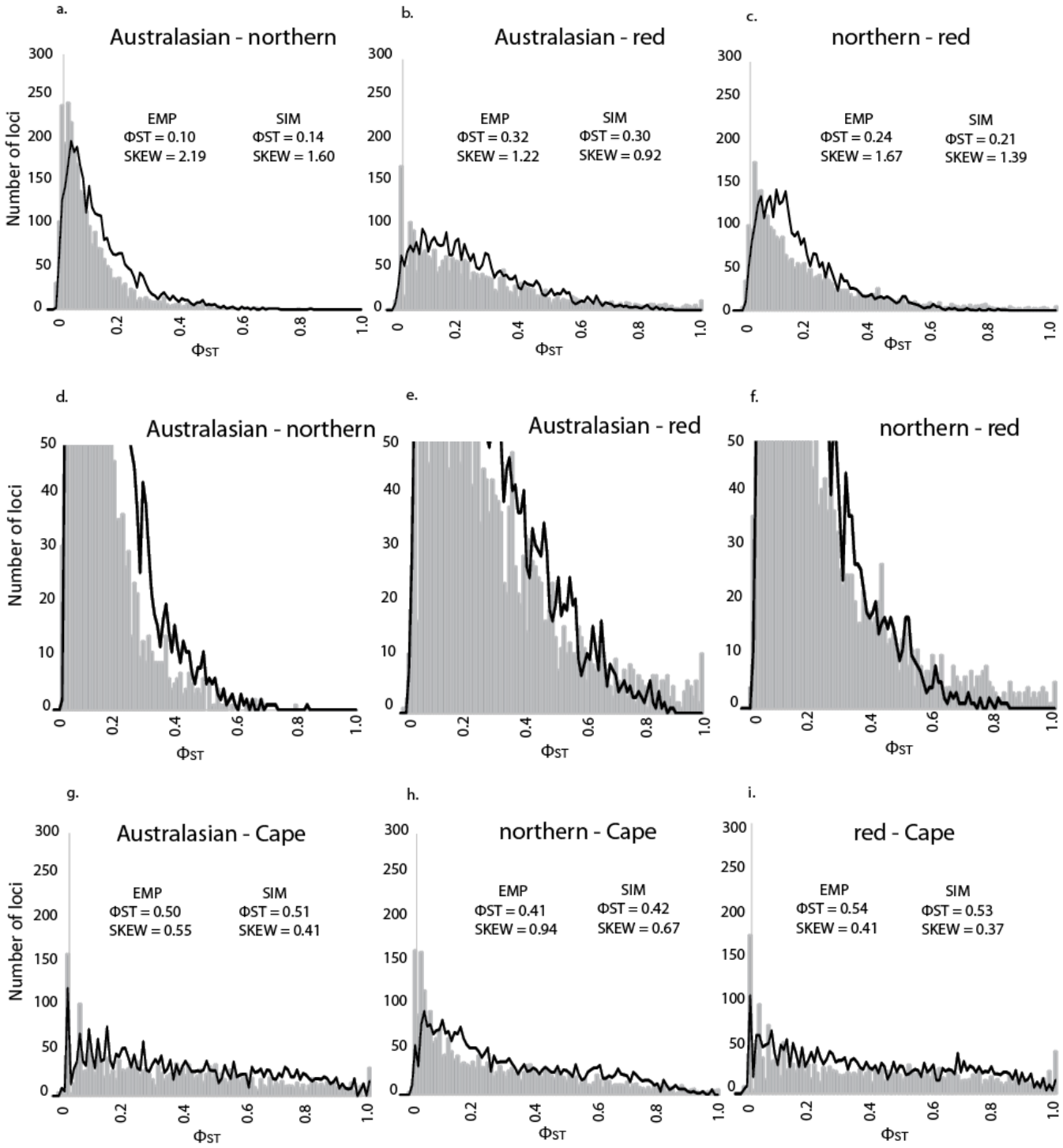

Figure 2-3: Histogram of ΦST values for all pairwise comparisons among shoveler species. Panels a-c and g-i show full Φ_{ST} distributions and panels d-f are zoomed-in versions of a-c to show the tails in the distributions. Empirical distributions, grey bars; simulated neutral distributions, black lines. Values within each panel give the overall Φ_{ST} and skewness (SKEW) for each empirical (EMP) and simulated (SIM) distribution.

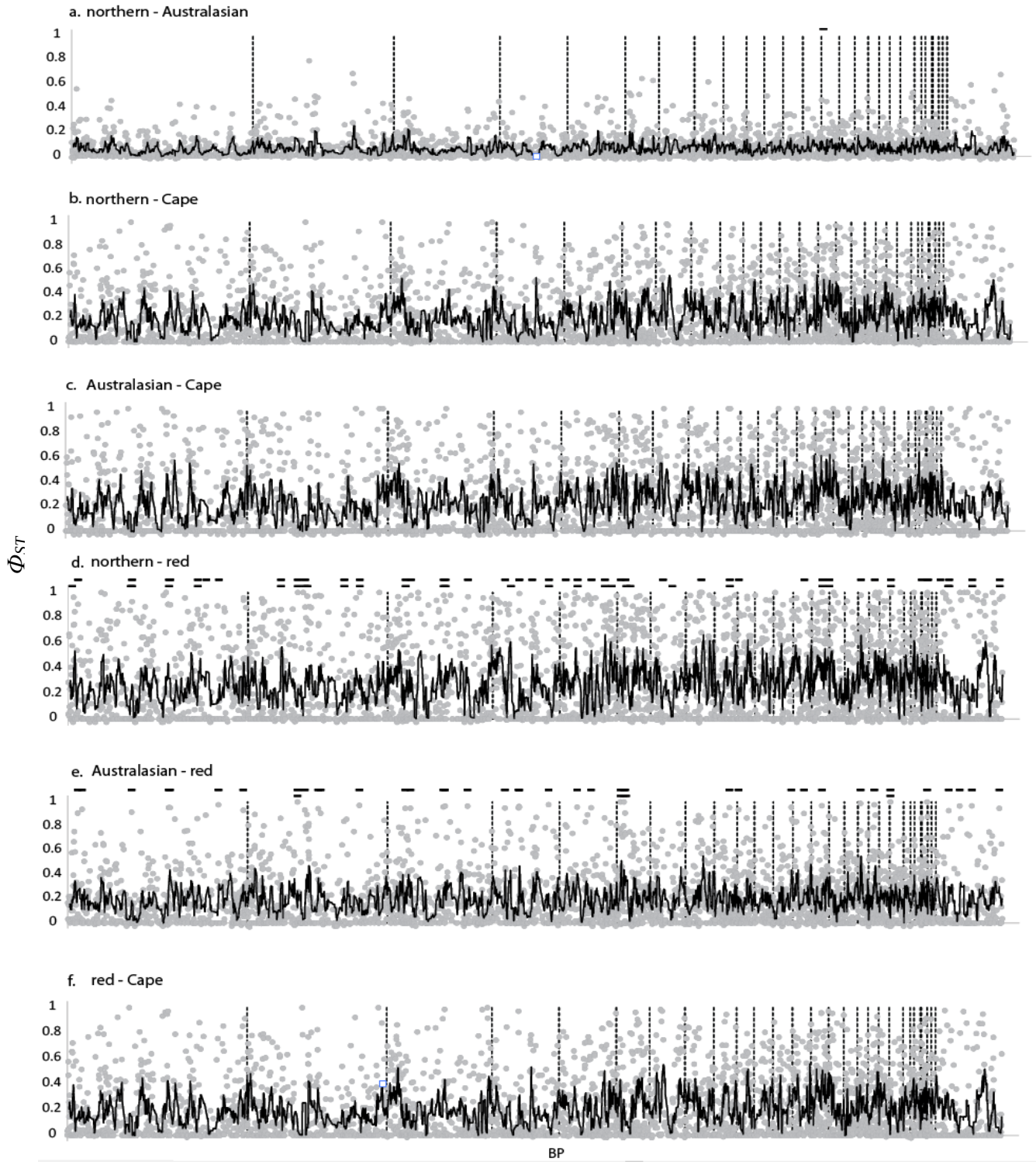

Figure. 2-4. Genomic scans for each pairwise comparison for all loci that were mapped to the mallard genome (N = 3,008 loci). Autosomes are ordered from one to twenty-eight and the Z-chromosome is on the right; chromosomes are separated by the vertical dashed lines. The black lines show the moving average of seven loci The dashes above each panel indicate outlier loci detected using BAYESCAN (bottom dashes) and neutral simulations (top dashes).